SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 14	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): NEARLY OPTIMAL DETECTION OF SIGNALS IN NON-GAUSSIAN NOISE	5. TYPE OF REPORT & PERIOD COVERED: Technical Report Sept. 1981-October 1983
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): S.V. Czarnecki and J.B. Thomas	8. CONTRACT OR GRANT NUMBER(s): N00014-81-K-0146
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Information Sciences and Systems Lab. Dept. of Electrical Eng. and Comp. Sci. Princeton Univ., Princeton, NJ 08544	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR SRO 103
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research (Code 411 SP) Department of the Navy Arlington, Virginia 22217	12. REPORT DATE: February 1984
	13. NUMBER OF PAGES: 214
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also submitted as the Ph.D. Thesis of S.V. Czarnecki to the EECS Department, Princeton University, Princeton, NJ 08544, October 1983

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Non-Gaussian Detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This dissertation addresses the problem of finding nearly optimal detector structures for non-Gaussian noise environments. It is assumed that the noise statistics are unknown except for a very loose characterization. Under this condition, the goal is to study adaptive detector structures that are simple, yet capable of high levels of performance.

Attention is focused on the discrete-time locally optimal

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

(con't.)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(Abstract con't.)

detector for a constant signal in independent, identically distributed noise. A definition for non-Gaussian noise is given, several common univariate density models are exhibited, and some physical non-Gaussian noise data is discussed.

Two approaches in designing adaptive detector nonlinearities are presented, where it is assumed that the noise statistics are approximately stationary. Both proposals utilize simple measurements of the noise behavior to adapt the detector, and in several examples the adaptive detectors are shown capable of attaining nearly optimal performance levels. A simulation is presented demonstrating their successful application.

The physical noise data is examined, and found to be contaminated with impulsive noise having a burst-like structure. This observation suggests that a nonstationary noise model and a time-varying detector may be appropriate. A nonparametric structure is proposed to detect the presence of impulsive bursts, and the performance of the detection algorithm is evaluated. It is then shown how information provided by the burst detector may be used to advantage in a signal detector. Performance of the combined detector structure is analyzed and found to be superior relative to the performance of any single fixed detector structure in certain noise environments. A simulation of the proposed structures is presented and compared to the simulation of the previous adaptive detectors.

The problem of approximating known locally optimal detector nonlinearities is examined and shown to be equivalent to the minimum mean square error approximation of known nonlinearities.

A performance index for comparing the performance of suboptimal threshold detectors operating with constant false alarm rates is proposed and analyzed. The ratio of indices for two detectors is shown to have appealing and useful properties in studying non-zero signal to noise ratio detection problems.

REPORT NUMBER 14

NEARLY OPTIMAL DETECTION OF SIGNALS IN NON-GAUSSIAN NOISE

S.V. CZARNECKI and J.B. THOMAS

INFORMATION SCIENCES AND SYSTEMS LABORATORY

Department of Electrical Engineering and Computer Science
Princeton University
Princeton, New Jersey 08544

FEBRUARY 1984

Prepared for

OFFICE OF NAVAL RESEARCH (Code 411SP)
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-81-K0146
SRO(103) Program in Non-Gaussian Signal Processing

S.C. Schwartz, Principal Investigator

NEARLY OPTIMAL DETECTION OF SIGNALS IN NON-GAUSSIAN NOISE

S.V. Czarnecki and J.B. Thomas
Department of Electrical Engineering
and Computer Science
Princeton University
Princeton, NJ 08544

ABSTRACT

This dissertation addresses the problem of finding nearly optimal detector structures for non-Gaussian noise environments. It is assumed that the noise statistics are unknown except for a very loose characterization. Under this condition, the goal is to study adaptive detector structures that are simple, yet capable of high levels of performance.

Attention is focused on the discrete-time locally optimal detector for a constant signal in independent, identically distributed noise. A definition for non-Gaussian noise is given, several common univariate density models are exhibited, and some physical non-Gaussian noise data is discussed.

Two approaches in designing adaptive detector nonlinearities are presented, where it is assumed that the noise statistics are approximately stationary. Both proposals utilize simple measurements of the noise behavior to adapt the detector, and in several examples the adaptive detectors are shown capable of attaining nearly optimal performance levels. A simulation is presented demonstrating their successful application.

The physical noise data is examined, and found to be contaminated with impulsive noise having a burst-like structure. This

observation suggests that a nonstationary noise model and a time-varying detector may be appropriate. A nonparametric structure is proposed to detect the presence of impulsive bursts, and the performance of the detection algorithm is evaluated. It is then shown how information provided by the burst detector may be used to advantage in a signal detector. Performance of the combined detector structure is analyzed and found to be superior relative to the performance of any single fixed detector structure in certain noise environments. A simulation of the proposed structures is presented and compared to the simulation of the previous adaptive detectors.

The problem of approximating known locally optimal detector nonlinearities is examined and shown to be equivalent to the minimum mean square error approximation of known nonlinearities.

A performance index for comparing the performance of suboptimal threshold detectors operating with constant false alarm rates is proposed and analyzed. The ratio of indices for two detectors is shown to have appealing and useful properties in studying non-zero signal to noise ratio detection problems.

Contents

1

Introduction

1. Motivation

Extracting information from raw data in the presence of noise is the ubiquitous problem of communication theory, and there are countless variations on this fundamental theme. In some contexts, it is important to estimate a signal or its parameters. In other contexts, it is desired to detect which, if any, signal is present. Both problems have received considerable theoretical and practical attention.

In this thesis, a very simple detection problem is posed: Detect the presence (or absence) of a known constant-level signal in a sequence of observations that is corrupted by addition of a sequence of observations from a random noise process. The problem is further simplified by assuming that the noise observations are statistically independent of each other and the signal. In several cases, it is further assumed that all noise observations are identically distributed.

In spite of this apparent simplicity, there remain some important issues: Namely, how does one approach the detection problem in non-Gaussian noise environments when the noise statistics are only partially known? What considerations are important in the design of a detection algorithm if the goal is to achieve a high level of performance with simple adaptive structures? How may a detector recognize an abrupt change in the noise and abate its effects? What properties should be possessed by a "good" procedure for approximating optimal detectors? How may the finite sample efficiency of a suboptimal detector be characterized?

2. Outline of the Thesis

This thesis has been written as one approach in addressing these and similar issues. The orientation of the work is not directed toward purely theoretical ends, nor is it purely an application of known results. Rather, it attempts to combine elements of both areas. The previous questions are studied in the combined light of abstract and practical considerations. Thus, the results which will be presented range from theorems and proofs to numerical simulations of proposed systems.

Chapter 2

Chapter 2 introduces the detection problem which is common to all chapters. Specifically, the Neyman-Pearson and locally optimal detectors are discussed, along with a description of their particular performance measures.

It is obvious that every density family, save for one, comprise *non-Gaussian densities*. These densities often characterize the noise in phy-

sical situations where classical assumptions leading to a Gaussian noise model are violated, and are of considerable practical interest. An implicit assumption in many cases is that the non-Gaussian densities deviate from the nominal Gaussian model in a particular way; most importantly, they are *heavy-tailed* relative to the Gaussian density. An explicit characterization for these densities is given in the chapter, and several useful non-Gaussian univariate density models are exhibited. The chapter concludes with a discussion of the importance of recognizing the effects of heavy-tailed noise and its impact on the detection problem as seen in previous work.

The appendix also introduces some non-Gaussian physical noise data which is used later in the thesis for simulation studies.

Chapter 3

The thrust of Chapter 3 is to consider the design of simple detector structures. It is assumed that little is known about the non-Gaussian noise environment, and that the goal is to design detectors with very simple structures and adaptation algorithms. Two alternative approaches are proposed: one is an "open loop" procedure where the observed noise density tails are characterized, and this information is used to update the detector structure. The other approach is a "closed loop" procedure where a very simple detector nonlinearity, a three-sectioned piecewise linear function, is proposed. An adaptive algorithm is then developed for finding the optimum nonlinearity parameters.

The performance of the two alternative structures and adaptation algorithms is examined under some known non-Gaussian noise density

models, as well as in a simulation using physical noise data to drive the algorithm.

Chapter 4

Chapter 4 also proposes a detector structure for a non-Gaussian noise environment, but assumes a different philosophy. There it is observed that stationary noise models may be inappropriate when the noise source contains bursts of impulsive noise; *i.e.*, the impulse producing event is short and well delineated in the noise observation sequence. Recognizing this fact, a nonstationary model for the noise is proposed, and a time varying detector structure is designed that capitalizes upon the ability of a subsidiary detector to recognize the presence of impulsive bursts. An algorithm for the subsidiary noise burst detector is developed using a nonparametric approach. The performance of the time varying detector and of the noise burst detector is examined in detail, and the physical noise data again is used to simulate the detection system.

Chapter 5

Answered in Chapter 5 is a question hinted at earlier in Chapter 3: what is the "best" way in which to approximate a known locally optimal detector structure? The term "best way" is interpreted as meaning the procedure yielding an approximation having the greatest efficacy, and a theorem is proven showing that the answer turns out to be any procedure that minimizes mean square error relative to the density induced measure. Implications of this theorem are discussed and its application is

illustrated in two examples.

Chapter 6

Throughout the thesis, concern is placed on locally optimal detection and the performance measure of efficacy. Chapter 6 changes course and examines the finite sample size performance of detectors which approximate the Neyman-Pearson structure. While closely related to previous work on performance bounding, this new work does not assume that the detector test statistic is generated by a likelihood ratio of the exact or approximate hypothesis densities. Bounds on the error probabilities are combined to form a single performance index, and several theorems establish its properties. The ratio of two indices is designated as the *relative bound efficiency*, which is shown to have a useful interpretation. The finite sample performance of some well known detectors is examined using relative bound efficiency.

Chapter 7

The results presented in the thesis are summarized in Chapter 7, and some suggestions for further study are made.

2

Signal Detection and the Non-Gaussian Noise Environment

Although diverse in purpose and form, radar, sonar, and data communication systems have at their heart a common important problem: detection of a signal in a noisy environment. This problem has received considerable attention in both the engineering and statistical literature, with viewpoints ranging from concrete details to abstract theory.

The purpose of this chapter is not to provide a thorough review of the detection problem, or of the noise environment modeling problem. Rather, this chapter is intended only to provide a common ground from which some particular problems in detection theory may be viewed; therefore mathematical rigor is suppressed for the sake of compactness. Complete exposition of the theory is available from the cited references. Section 1 provides a short introduction to the detection problem and the

theoretical foundations upon which the remainder of the thesis will rest. Specifically, the Neyman-Pearson and locally optimal detector structures are introduced with their associated performance measures. In Section 2, note is made of a particular type of noise environment which will be of concern in this thesis, and some noise models are discussed. The notion of a *non-Gaussian density* is developed to the degree necessary to give it a characterization. Finally, Section 3 discusses the impact of non-Gaussian noise on the detection problem and summarizes some results which are background and motivation to the approaches in this thesis. The Appendix outlines the characteristics of some physical noise data which is used later in the thesis to drive various simulations.

1. Detector Structures and Performance Measures

Neyman-Pearson Detector Structure

A binary hypothesis test may be used to model the problem of detecting a known signal in the presence of noise. Consider the following detection problem in discrete time over a signaling interval of length M. Let $\theta\mathbf{s} = \theta\{s_1, \ldots, s_M\}$ be a known signal sequence with amplitude parameter $\theta > 0$, and let $\mathbf{n} = \{n_1, \ldots, n_M\}$ be an independent identically distributed (iid) noise sequence independent of the signal. Section 2 will provide justification for the iid restriction on the noise. The detector observes $\mathbf{x}$, a data sequence $\{x_1, \ldots, x_M\}$, and decides between:

$$H_0: \mathbf{x} = \mathbf{n}$$
$$H_1: \mathbf{x} = \mathbf{n} + \theta\mathbf{s}$$

Here, without loss of generality, we restrict ourselves to the special case

of distinguishing between the two signals $\mathbf{s}_0 \equiv 0$ and $\mathbf{s}_1 \equiv \mathbf{s}$. In the framework of Neyman-Pearson (NP) hypothesis testing [1-3], the observation $\mathbf{x}$ and the multivariate noise density $f_{\mathbf{N}}$ are used to calculate a likelihood ratio Λ_{NP}. This test statistic and a fixed threshold T_{NP} are compared to arrive at a decision: H_1 is chosen when $\Lambda_{NP} > T_{NP}$, and H_0 is chosen when $\Lambda_{NP} \leq T_{NP}$. More precisely,

$$\Lambda_{NP} = \frac{f_{H_1}(\mathbf{x})}{f_{H_0}(\mathbf{x})} = \frac{f_{\mathbf{N}}(\mathbf{x}-\theta\mathbf{s})}{f_{\mathbf{N}}(\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} T_{NP} \tag{2.1}$$

Since the noise is iid, $f_{\mathbf{N}}(\mathbf{n}) = \prod_{i=1}^{M} f(n_i)$ where f is the univariate density of the noise. *For the sake of brevity, in the remainder of this thesis we adopt the convention that $f(\cdot)$ is the univariate noise density unless explicitly stated otherwise.* Because the logarithm function is monotonic a test equivalent to (2.1) is

$$\lambda_{NP} = \ln \Lambda_{NP} \underset{H_0}{\overset{H_1}{\gtrless}} t_{NP} = \ln T_{NP} \tag{2.2}$$

where

$$\lambda_{NP} = \sum_{i=1}^{M} g_{NP;i}(x_i) = \sum_{i=1}^{M} \ln \frac{f(x_i - \theta s_i)}{f(x_i)} \tag{2.3}$$

The function $g_{NP;i}$ is memoryless, but time varying because the signal varies with time. Consideration of the time varying signal case adds nothing beyond unnecessary complication to the essence of this discussion. Therefore, we will sacrifice completeness for clarity, restrict attention to the constant signal $s_i = s$ for $i = 1, \ldots, M$, and replace the sequence

$\{g_{NP,i}\}$ with a memoryless nonlinearity, g_{NP}. Figure 2.1 presents a block diagram of the NP detector structure generated by (2.2) and (2.3).

Neyman-Pearson Detector Performance

The performance of a Neyman-Pearson detector is usually measured in terms of its *false alarm rate* α and its *power* of detection β. These quantities are defined as

$$\alpha = \text{Prob}(\text{say } H_1 \,|\, H_0 \text{ true})$$

$$\beta = \text{Prob}(\text{say } H_1 \,|\, H_1 \text{ true})$$

These probabilities are determined by the distribution of λ_{NP} under H_0 and H_1, respectively, and the value of the threshold t_{NP}. Thus,

$$\alpha = \int_{t_{NP}}^{\infty} p_{H_0}(\lambda) d\lambda \tag{2.4}$$

$$\beta = \int_{t_{NP}}^{\infty} p_{H_1}(\lambda) d\lambda \tag{2.5}$$

The Neyman-Pearson detector is optimal [1-3] in the sense that, for any given false alarm rate α_0, the NP test achieves the maximum probability of detection β in the set of all possible tests with $\alpha \leq \alpha_0$.

The performance measures α and β are not restricted to the characterization of NP tests only. The performance of any threshold detection scheme may be parameterized via (2.4) and (2.5). In principle, if the noise density f is known, and a known nonlinearity g processes the observations, then p_{H_0} and p_{H_1} may be computed via transformation of the hypothesis densities in the case where $\mathbf{x}$ is a single observation and

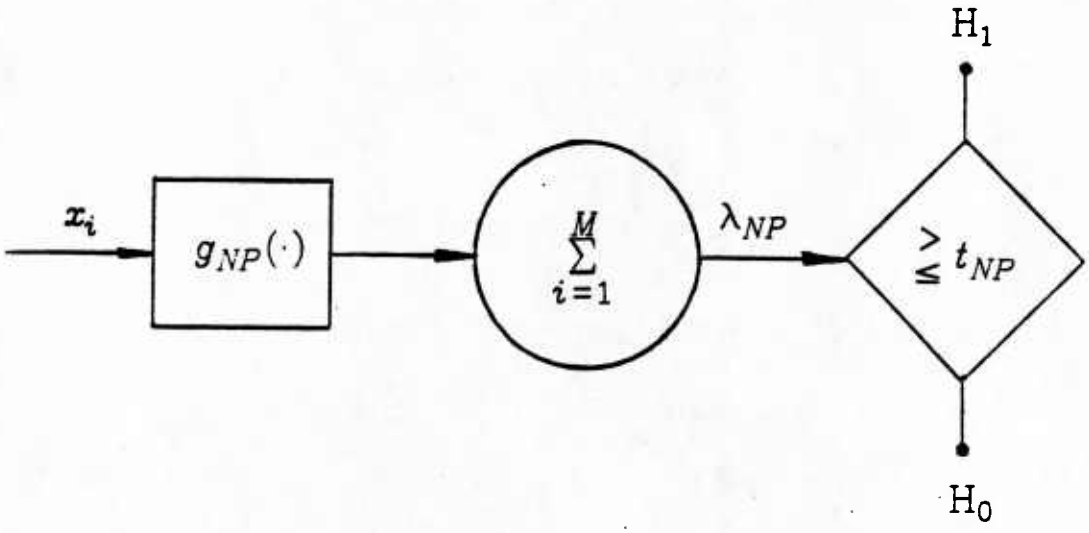

Fig. 2.1. Block diagram of the Neyman-Pearson optimal detector structure for detection of a known constant signal in additive iid noise.

$M = 1$. For multiple observations, $M > 1$, and p_{H_0} and p_{H_1} may be computed via M-fold convolutions of the transformed hypothesis densities, as is well known. See, for example, references [3 pp. 215-219] and [4]. In other instances, carrying out this procedure may be a difficult or intractable problem, especially when M is large. One must then resort to calculating α and β via Monte Carlo simulations [5,6], series expansions [3, pp. 219-226] and [7], numerical approximation methods [8,9,49], performance bounding [2, pp. 116-133] and [10], or the Central Limit Theorem [11 pp. 308-319] and [4].

The measures α and β are often inconvenient to compute, even though they give a complete characterization of detector performance. Small changes in the detector nonlinearity or noise density may change a tractable computation into an intractable problem. Further, with the exception of Central Limit Theorem based techniques, most methods give little qualitative or quantitative insight into understanding how changes in the threshold or sample size affect performance. All of the techniques offer little illumination of performance sensitivity as a function of changes in the nonlinearity shape.

Locally Optimal Detector Structure

In cases where the signal-to-noise ratio is very small, $\theta \approx 0$, and the test statistic may be calculated via the locally optimal (LO) detector [12,13]. The test becomes

$$\lambda_{LO} = \sum_{i=1}^{M} g_{LO;i}(x_i) \underset{H_0}{\overset{H_1}{\gtrless}} t_{LO} \tag{2.6}$$

where

$$g_{LO;i}(x_i) = \frac{d}{d\theta} \ln \frac{f(x_i - \theta s_i)}{f(x_i)} \bigg|_{\theta=0} = -\frac{f'(x_i)}{f(x_i)} s_i \qquad (2.7)$$

Simply put, $g_{LO;i}$ is the coefficient of θ in the Taylor series expansion of $g_{NP;i}$ about $\theta = 0$. Despite the fact that $\mathbf{s}$ may be time varying, the transformation operating on x_i is not a function of i. Instead, (2.6) and (2.7) imply that the output of a single memoryless nonlinearity g_{LO} should be correlated with the signal. Once again, to simplify the discussion, we restrict the signal to be constant, rescale the test statistic by s, and limit our efforts to consideration of $g_{LO} = -f'/f$. If g_{LO} is substituted for g_{NP}, then Figure 2.1 also represents a block diagram of the LO detector structure generated by (2.6) and (2.7).

Locally Optimal Detector Performance

Rather than derive rigorously the performance measures of efficacy and asymptotic relative efficiency for locally optimal detectors, we briefly summarize some of the important points of these useful measures. A thorough treatment of this subject is available in [3,12-14].

The Neyman-Pearson detector is optimal in the sense that for given α, it maximizes β when the signal amplitude θ is nonzero. The locally optimal detector, on the other hand, is optimal in the sense that, for given α, it maximizes $\frac{\partial}{\partial\theta}\beta(\theta)\Big|_{\theta=0}$. Zero signal strength is obviously a limiting worst case. A useful way of comparing two detectors in this limiting case is to compute their ARE, or asymptotic relative efficiency, where,

$$\mathrm{ARE}_{g_1,g_2} = \lim_{\theta \to 0} \frac{M_2(\alpha,\beta,\theta)}{M_1(\alpha,\beta,\theta)} \tag{2.8}$$

and $M(\alpha,\beta,\theta)$ is interpreted as the number of data observations necessary to provide a (α,β) performance level for signal level θ. As seen in (2.8), asymptotic relative efficiency is the ratio of the number of samples necessary in each of two alternative detectors to maintain the same probabilities of false alarm and detection as the signal-to-noise ratio approaches zero. Regularity conditions [13] ensure that as $\theta \to 0$, both M_1 and $M_2 \to \infty$; thus, ARE is a small signal, large sample size measure of performance.

A simple expression is available to compute ARE, and is defined as

$$\mathrm{ARE}_{g_1,g_2} = \frac{\eta_f(g_1)}{\eta_f(g_2)} \tag{2.9}$$

where $\eta_f(g)$ is the *efficacy* of detector g in noise density f. It may be shown [3, p. 228; 13], for iid noise and a fixed detector nonlinearity g, that

$$\eta_f(g) = \frac{\left[\int_{-\infty}^{\infty} g'(x) f(x) dx\right]^2}{\int_{-\infty}^{\infty} g^2(x) f(x) dx} \tag{2.10}$$

where g has zero mean under f. This definition of efficacy is subject to the following regularity conditions in a neighborhood of $\theta = 0$:

(*i*) $\dfrac{\lambda_g - \mathrm{E}_{H_1}\lambda_g}{\mathrm{var}_{H_0}\lambda_g}$ is asymptotically normal with zero mean and unity variance. Here, λ_g is the test statistic of a detector using

memoryless nonlinearity g.

(*ii*) $\theta = kM^{-\frac{1}{2}}$, where k is a nonzero constant.

(*iii*) $\lim_{\theta \to 0} \frac{\partial}{\partial \theta} E_{H_1} g = E_{H_0} g'$

The LO detector maximizes efficacy [13]; as a result $ARE_{g,g_{LO}} \leq 1$ for all detectors g satisfying the regularity conditions. Note that (2.10) is a ratio of two expectations, and is therefore usually more convenient to compute than an M-fold convolution of a probability density.

A criticism of ARE is that it is a performance measure based upon a limiting case. While ARE measures asymptotic performance, it may not give a good indication of the relative merits of two detectors in a small sample, nonzero signal environment. Some recent work [15,16,50] has concentrated on examining the convergence of relative efficiency to ARE. Further, because efficacy is only a ratio of two expectations, it may be argued that two detectors with identical efficacies need not have similar small sample performance.

Despite these criticisms, ARE, efficacy, and the LO detector shall receive the most attention in this thesis for several reasons. First, efficacy is a convenient, accepted, and well studied measure of performance. Second, small signal detection is an important problem, and the zero (or infinitesimal) signal is the limiting bound. The NP and LO detectors are asymptotically equivalent in the limiting case, even though LO detection is not optimal for a nonzero signal [13,51]. Third, for very small signals there is a close correspondence between the forms of the NP and LO detectors, which suggests that by paying attention to the LO detection

problem, it should be possible to gain insight into the issue of NP detector design. The relationship between the nonlinearities for the two types of detectors is

$$sg_{LO}(x) = \frac{\partial}{\partial \theta} \ln g_{NP}(x) \Big|_{\theta=0} \tag{2.11}$$

and, from [17]

$$g_{NP}(x) = \int_{x-\theta s}^{x} g_{LO}(x)dx \tag{2.12}$$

The latter equation implies that if θ is small, and if g_{LO} is approximately constant over the range of integration, then $g_{NP} \approx \theta s g_{LO}$.

2. The Non-Gaussian Noise Environment

General Assumptions

Before discussing some noise models of interest later in the thesis, it is necessary to state some of the fundamental assumptions about the noise environment and the models which will be used. First, as stated in the previous section, we are interested in the discrete time environment. Second, we assume that the noise samples are independent and identically distributed. This is a very strong assumption with extensive and rigorous requirements on the noise behavior, but it allows simplification of the analysis. For example, the difficult problem of modeling dependent non-Gaussian noise is eradicated by the independence assumption. Further, a noise with a nonstationary distribution implies that time varying detector structures are necessary, which, in general, may be quite difficult to specify and implement.

What is of more interest than strict mathematical fulfillment of the iid requirement is that the assumption be *approximately* true for the physical case of interest. Even though noise is usually correlated due to the finite bandwidth of a channel, adjacent samples may be approximately independent, provided the sampling rate is low enough. The noise environment is always nonstationary, as no real source has unchanging statistics for the infinite past and the infinite future. Over finite intervals, however, the statistics may appear stationary, or the noise statistics may be changing slowly enough that they appear approximately stationary and may be tracked by an adaptive system.

To provide a starting point, then, it is not unreasonable to assume iid noise. This assumption is a divergence from the reality of physical noise environments, but for that price clarity and mathematical simplicity are purchased. An implication of this assumption is that the noise environment is described adequately by a univariate density.

There is an abundance of information on the measured statistics of physical noise sources, a full report on which is beyond the intended scope of this chapter. Instead, as the emphasis is on understanding the problem of finding near-optimal detectors for non-Gaussian noise, the following subsections present some common noise models which will be used in the following chapters. *For convenience, the noise densities will be assumed here to be zero mean and unit variance.*

Gaussian Noise Model

A Gaussian noise background is the classical assumption in the design and analysis of detection systems. Here, the univariate noise

density is the well known expression

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{2.13}$$

which leads to the LO nonlinearity

$$g_{LO}(x) = x \tag{2.14}$$

For convenience, (2.14) will be referred to as the linear detector, *ld*.

The Gaussian assumption has attractive features in that it is mathematically tractable and the optimal detector structure is a linear processor. Strong justification for this noise model is available due to Central Limit Theorem (CLT) arguments, for at least two reasons: first, the noise source often may be considered as a shot noise process comprising a very large number of small effects with additive cumulative effect; *e.g.*, thermal noise. Second, the finite bandwidth of many channels "averages" together the noise process, tending to make the noise at the channel output Gaussian. In the limit, as the channel bandwidth approaches zero, it may be shown [18 Thm. 2.4] that the noise output process of a narrowband channel converges in distribution to a Gaussian process.

Rebuttal of the Gaussian Model

Despite these arguments, measurements of different noise environments have led to the conclusion that the true noise distribution often is described better by a heavier tailed pdf; *e.g.* [19-22, 25-28,30,35,36,48]. Also, see the discussion and bibliographies of [17,23,32,33]. This type of noise may be ascribed to a nominal Gaussian environment with a heavy

tailed impulsive noise contaminant. Another consideration is that a real noise comprises a finite sum of random events; in a shot process there can only be a finite number of contributors, and any real channel has nonzero bandwidth. The result is that CLT convergence to the Gaussian pdf is not complete. Instead, the observed noise pdf is most nearly Gaussian near the mean, with the tails converging to the Gaussian pdf only in the limiting case.

For example [24 p. 103], if $X_m = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_i$, and the Y_i are iid random variables with continuous distribution function, $EY = 0$, $EY^2 = 1$, and $EY^3 = 0$, then $\Phi(x_m) \approx F_{X_m}(x_m)$ for $|x_m| \leq \sqrt{\ln m}$ where Φ is the Gaussian distribution function. Convergence of the sum distribution to the Gaussian is from the mean outwards, and the size of the Gaussian-like region is proportional to $\sqrt{\ln m}$.

Contained in this discussion is a partially constructive, but loose, characterization of *non-Gaussian noise densities*. Obviously, all families of densities save for one are non-Gaussian. However, in this thesis only particular types of deviations from the nominal Gaussian family are of concern. The term non-Gaussian noise density will refer to unimodal, symmetric densities which have a Gaussian-like shape in a region centered about the mode. These densities also possess tails that are heavier than the Gaussian, for they converge to zero asymptotically at a rate less than an equal variance Gaussian density. This type of density is often referred to as being *heavy-tailed* or *long-tailed*.

Middleton's Class A and B Density Models

Noise density models may be classified into one of two categories: physically motivated models and empirical models. The first group of models take into consideration physical aspects of the noise situation and attempt to describe the density from this physical accounting. The second group of models use convenient distributions which seem to agree well with observed characteristics of the noise.

Middleton's Class A and Class B models fall into the category of physically motivated models. Without exposition of the details found in [20,25-28], both models intend to characterize situations where the noise is nominally Gaussian with an additive impulsive noise component. The Class A model assumes that these spikes are of lesser bandwidth than the receiver, and as such, do not generate a transient response of significant duration relative to the spike duration. The Class B model assumes the reverse, and the spikes produce relatively long transients.

The Class B model comprises an infinite series of confluent hypergeometric functions, each of which is generally defined by an infinite series [29 p. 504]. Because of its unwieldiness, we will not consider it further.

The Class A Model comprises an infinite series of Gaussian density-like terms, and may be written as

$$f_A(x) = \sum_{m=0}^{\infty} \frac{e^{-A} A^m e^{-x^2/2\sigma_m^2}}{m!} \qquad (2.15)$$

The parameter A is called the overlap index and is the product of the duration of individual events in the impulsive component and the mean

rate of the shot process generating the impulsive component events. The other parameter is defined as

$$\sigma_m^2 = \frac{m/A+\Gamma'}{1+\Gamma'}$$

where Γ' is a measure of the ratio of the power in the impulsive component compared to the power in the Gaussian background. Both parameters are directly related to physical measurements of the noise environment [30].

Figures 2.2 - 2.5 compare some representative unit variance Class A densities and the Gaussian density. The Class A densities have Gaussian-like behavior near $x=0$, as evidenced by their parabolic shape on the log-scaled plots. For large x, however, they have a much heavier tail behavior than the Gaussian density.

The LO nonlinearity associated with f_A may be written as

$$g_A(x) = x \left(\frac{\sum_{m=0}^{\infty} \frac{A^m e^{-x^2/2\sigma_m^2}}{\sigma_m^2 m!}}{\sum_{m=0}^{\infty} \frac{A^m e^{-x^2/2\sigma_m^2}}{m!}} \right) \tag{2.16}$$

Figures 2.6 and 2.7 compare the LO nonlinearities for a Gaussian density and the Class A densities of the previous figures. While g_A has nearly linear behavior for small $|x|$, the effect of large observations is greatly reduced with respect to a linear processor.

The Gaussian-Gaussian ε-mixture Family

Another useful interesting class of noise distributions is the Gaussian-Gaussian ε-mixture family. It may be written as

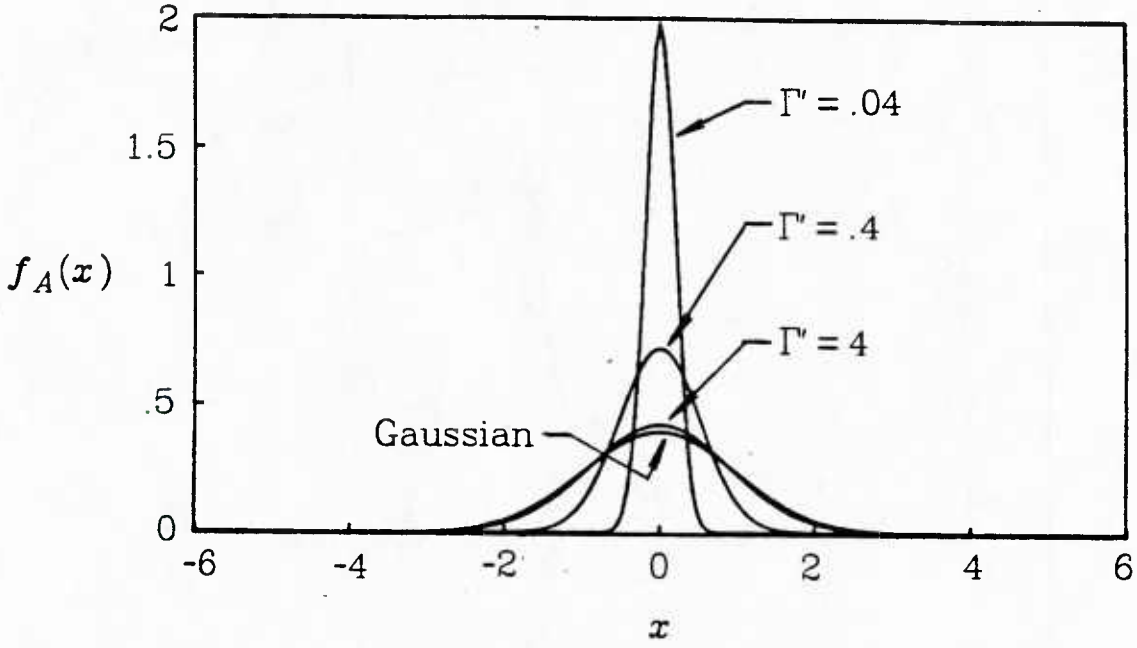

Fig. 2.2. Representative Middleton Class A densities f_A with $A = .05$ compared to Gaussian density.

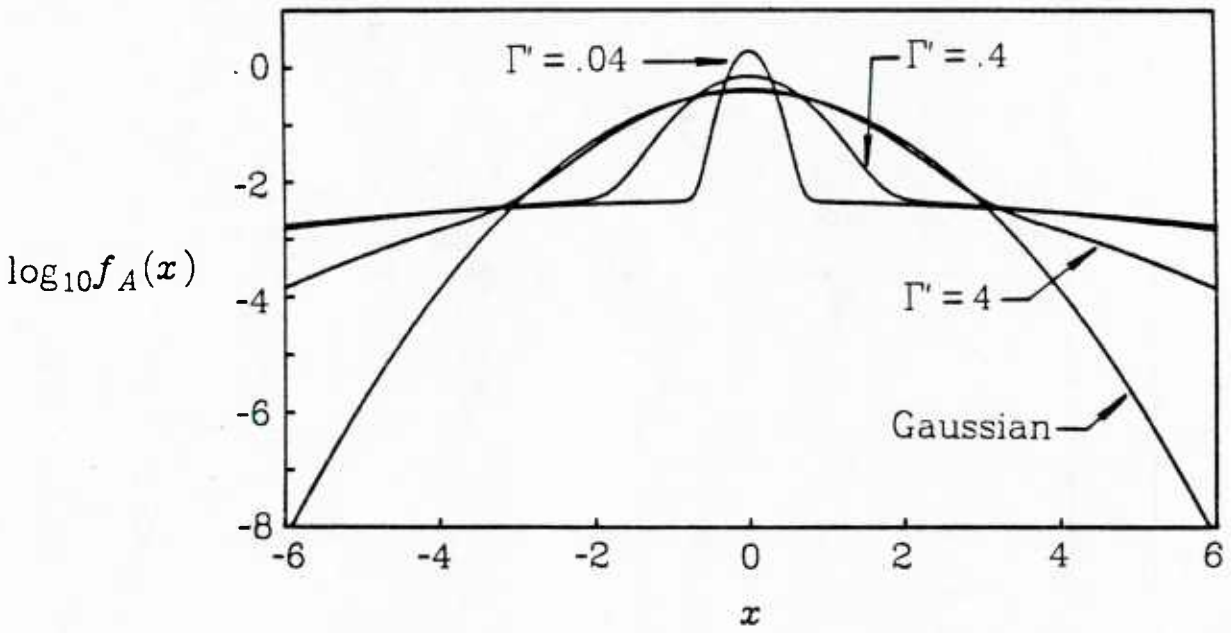

Fig. 2.3. Logarithm of Middleton Class A densities f_A with $A = .05$ compared to log of Gaussian density.

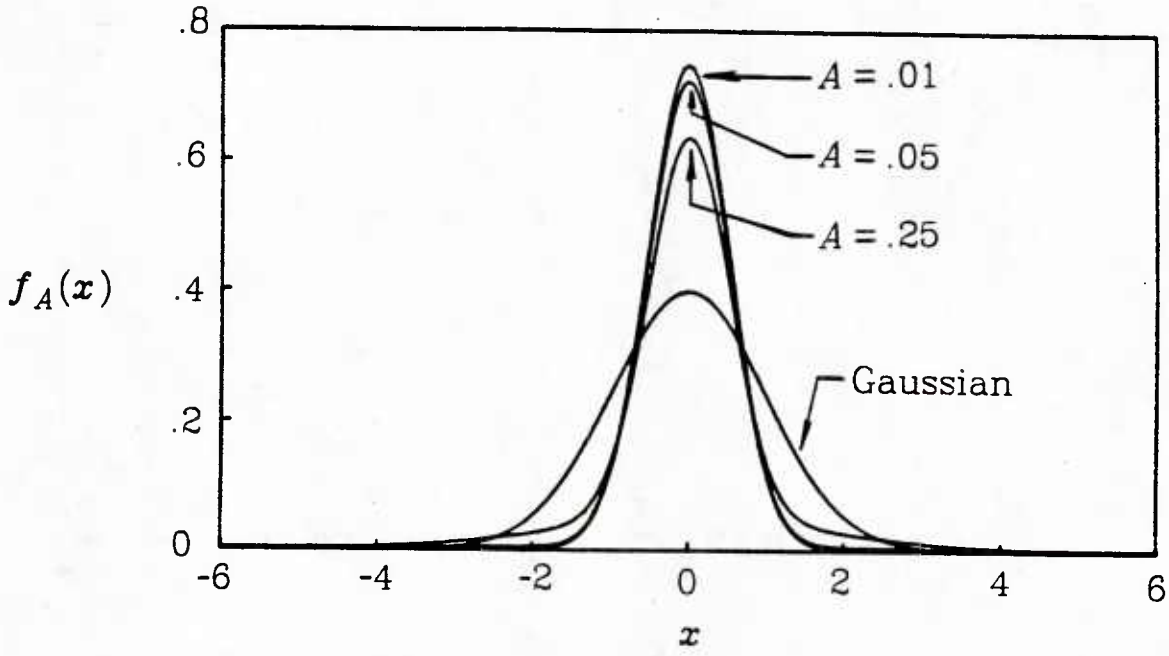

Fig. 2.4. Representative Middleton Class A densities f_A with $\Gamma' = .4$ compared to Gaussian density.

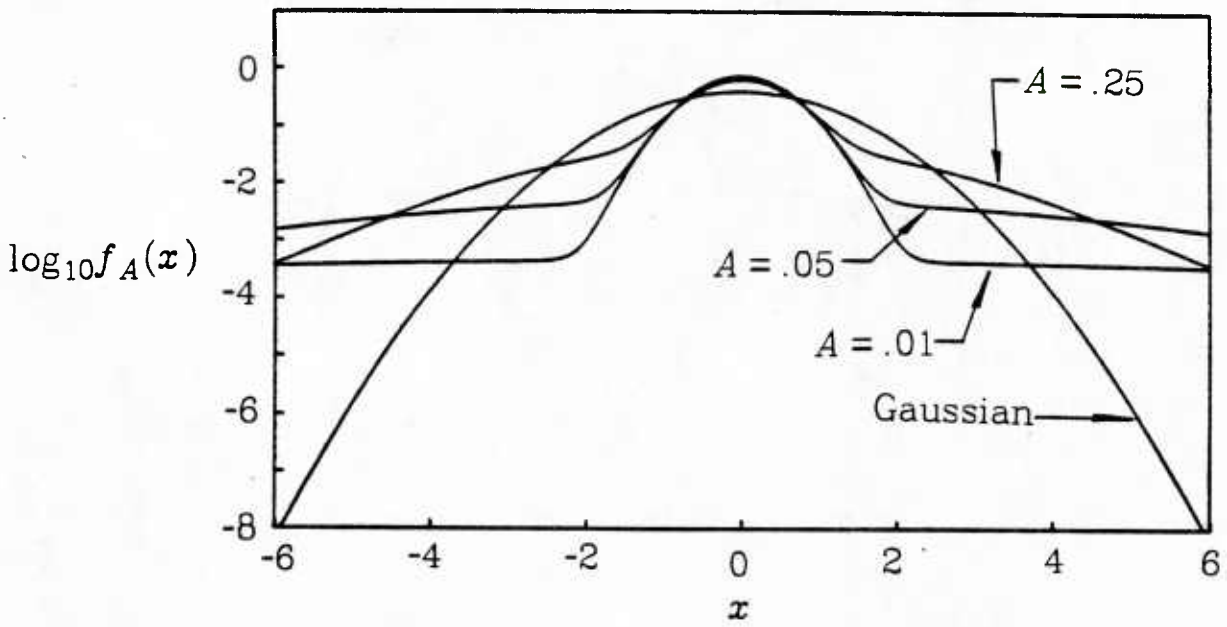

Fig. 2.5. Logarithm of Middleton Class A densities f_A with $\Gamma' = .4$ compared to log of Gaussian density.

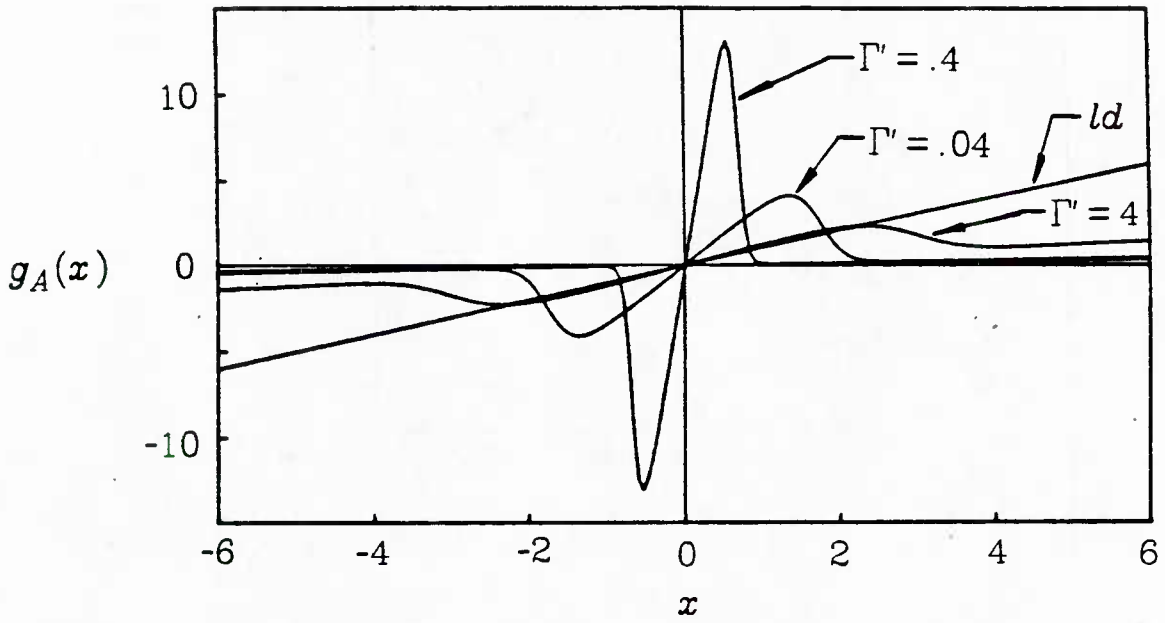

Fig. 2.6. Locally optimal nonlinearity g_A for Middleton Class A densities with $A = .05$ compared to unity slope linear detector *ld*.

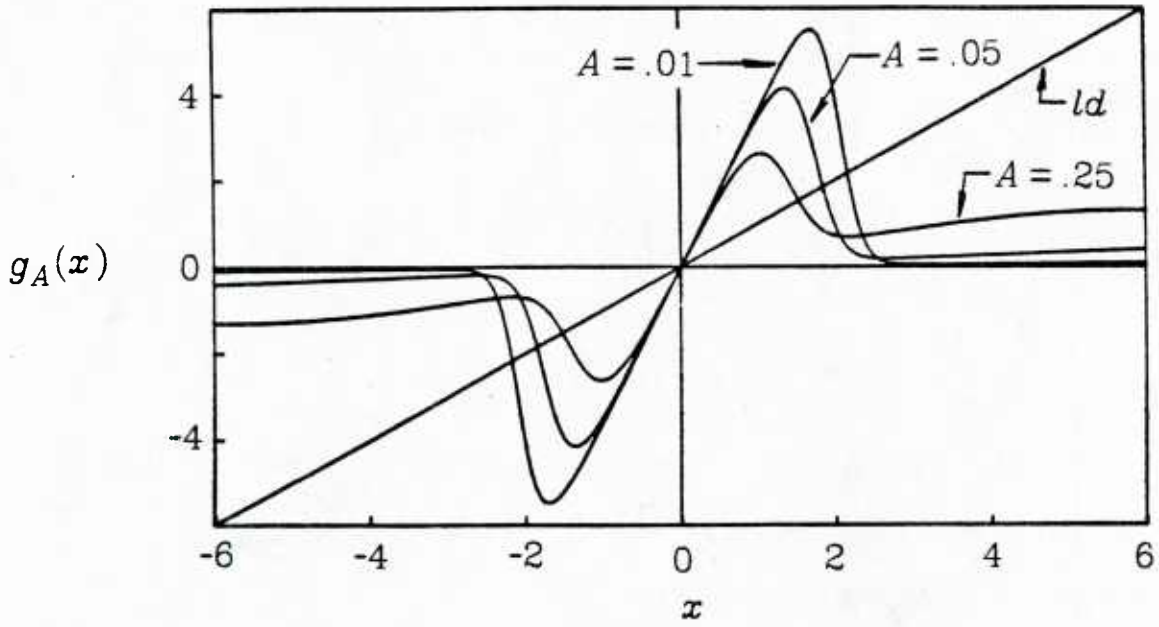

Fig. 2.7. Locally optimal nonlinearity g_A for Middleton Class A densities with $\Gamma' = .4$ compared to unity slope linear detector *ld*.

$$f_\varepsilon(x) = (1-\varepsilon)f_0(x)+\varepsilon f_1(x) \tag{2.17}$$

where f_0 and f_1 are both zero mean Gaussian densities, with $0 \le \varepsilon < 1$ typically assuming a small value and $\sigma_1^2 > \sigma_0^2$.

The LO nonlinearity associated with this density is

$$g_\varepsilon(x) = x\left[\frac{\frac{(1-\varepsilon)}{\sigma_0^2}f_0(x) + \frac{\varepsilon}{\sigma_1^2}f_1(x)}{(1-\varepsilon)f_0(x)+\varepsilon f_1(x)}\right] \tag{2.18}$$

Figures. 2.8 - 2.11 compare some representative Gaussian-Gaussian ε-mixture densities and the Gaussian density. Figures 2.12 and 2.13 compare the LO nonlinearities g_ε to a linear processor.

The density f_ε is attractive in that it is a relatively simple empirical model, and has been proposed for describing heavy tailed non-Gaussian noise [31,32]. Recently [27,33], it was shown that it also may be considered as a tractable simplification of Middleton's Class A Model arising by truncating all terms of f_A for $m > 1$. The parameters of f_ε have a simple relationship to the parameters of f_A, given here as

$$\varepsilon = \frac{A}{1+A} \tag{2.19}$$

$$\frac{\sigma_1^2}{\sigma_0^2} = 1+\frac{1}{A\Gamma} \tag{2.20}$$

Therefore, f_ε may be considered a quasi-physically based model. Figure 2.14 compares the LO nonlinearity for f_A and the corresponding LO nonlinearity for the approximating density f_ε.

In addition to a Gaussian-Gaussian mixture, others have considered

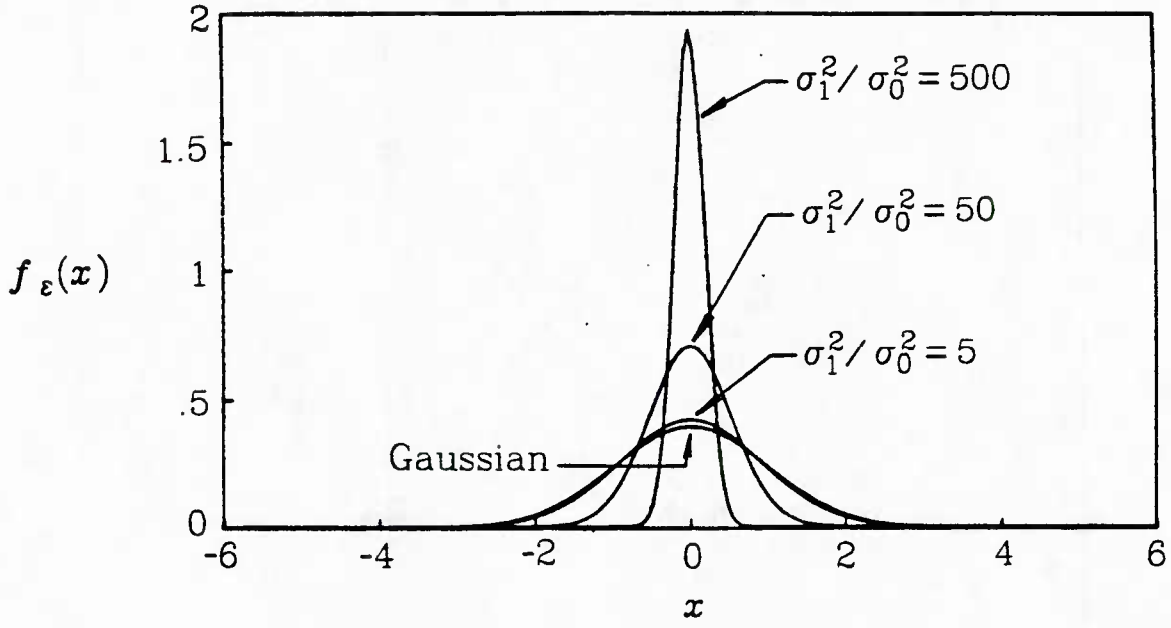

Fig. 2.8. Representative Gaussian-Gaussian ε-mixture densities f_ε with $\varepsilon = .05$ compared to Gaussian density.

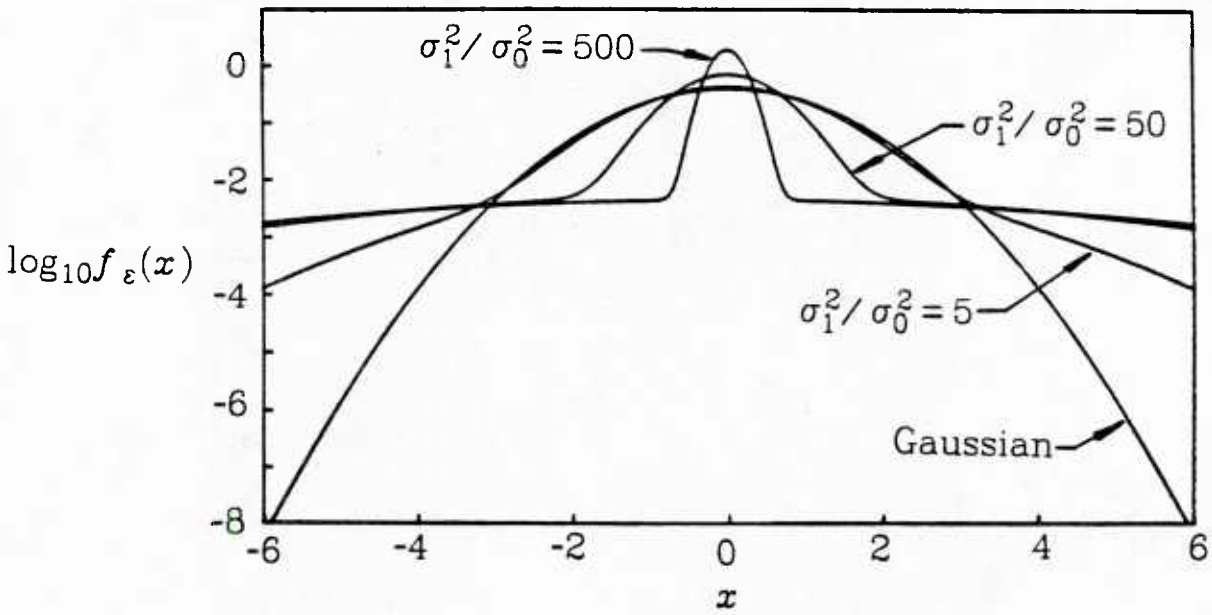

Fig. 2.9. Logarithm of Gaussian-Gaussian ε-mixture densities with $\varepsilon = .05$ compared to log of Gaussian density.

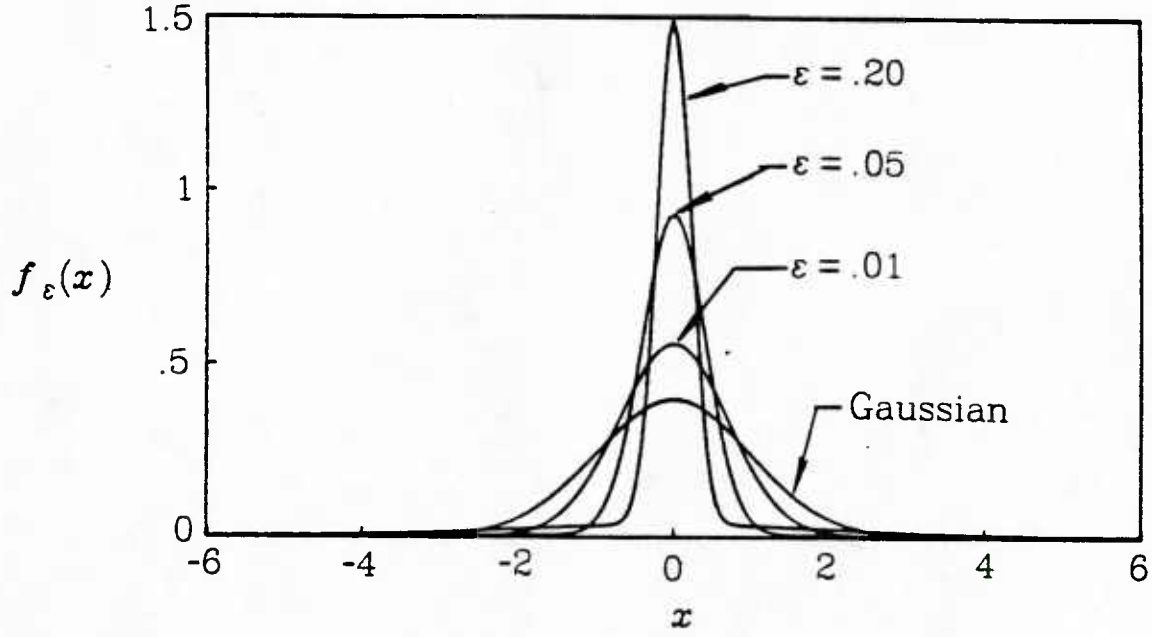

Fig. 2.10. Representative Gaussian-Gaussian ε-mixture densities with $\sigma_1^2/\sigma_0^2 = 100$ compared to Gaussian density.

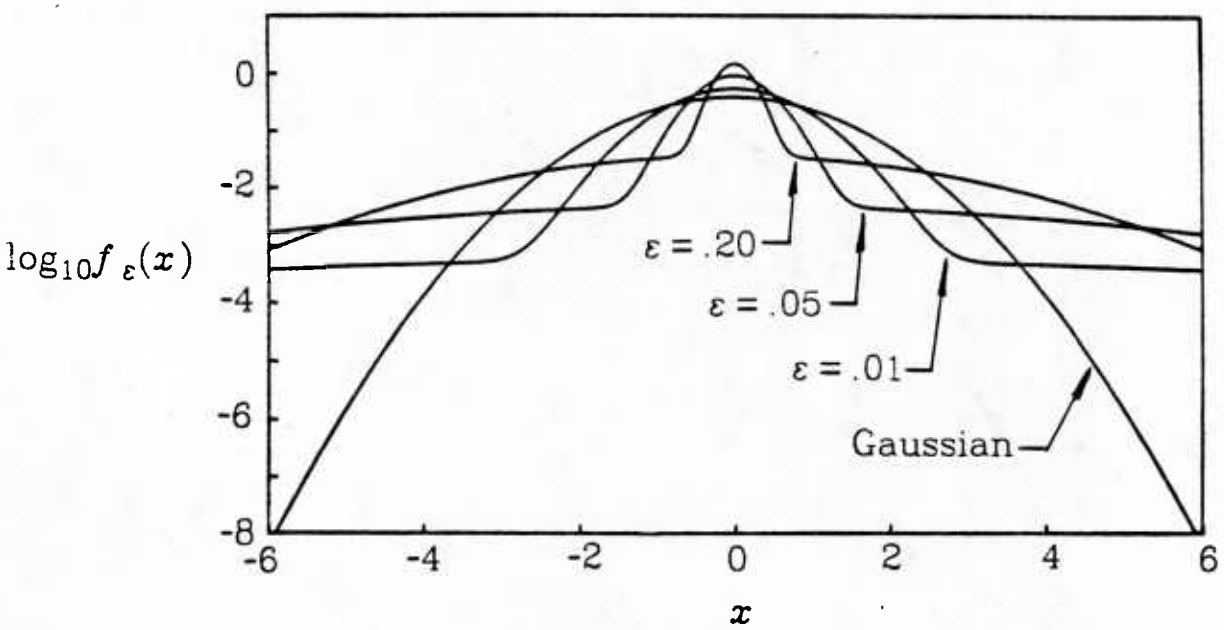

Fig. 2.11. Logarithm of Gaussian-Gaussian ε-mixture densities with $\sigma_1^2/\sigma_0^2 = 100$ compared to log of Gaussian density.

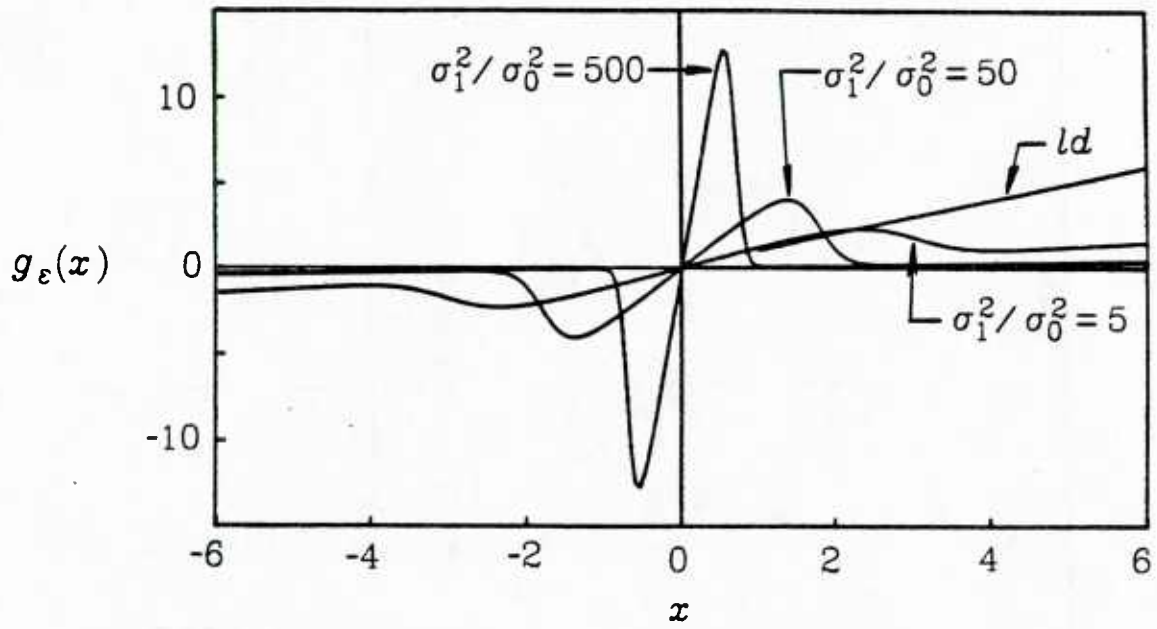

Fig. 2.12. Locally optimal nonlinearity g_ε for Gaussian-Gaussian ε-mixture densities with $\varepsilon = .05$ compared to unity slope linear detector ld.

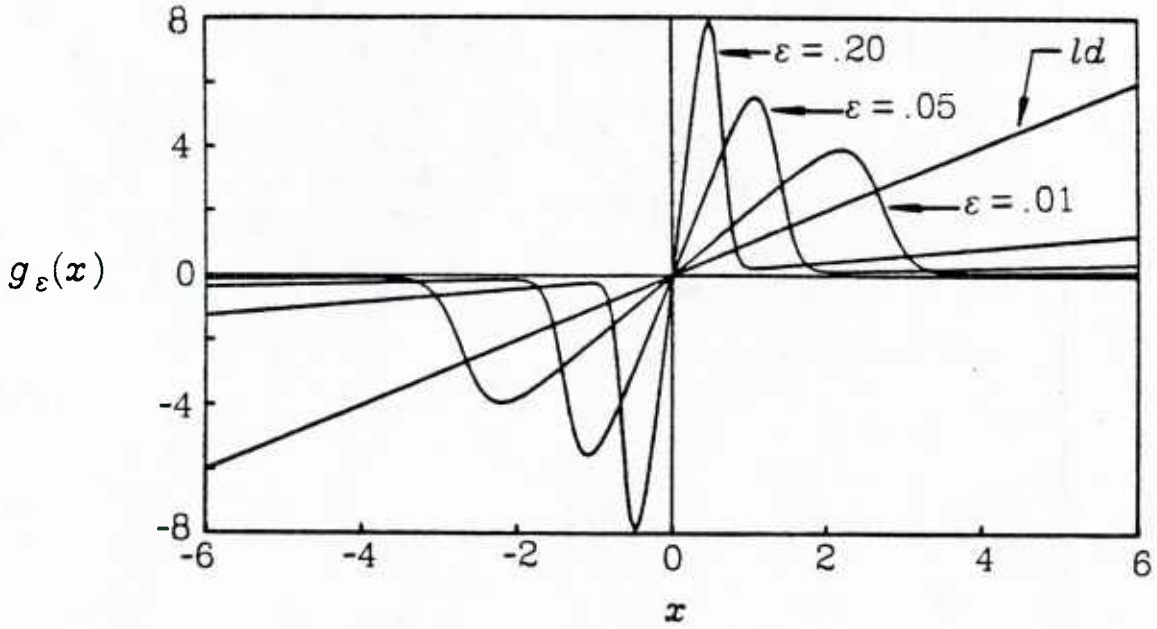

Fig. 2.13. Locally optimal nonlinearity g_ε for Gaussian-Gaussian ε-mixture densities with $\sigma_1^2/\sigma_0^2 = 100$ compared to unity slope linear detector ld.

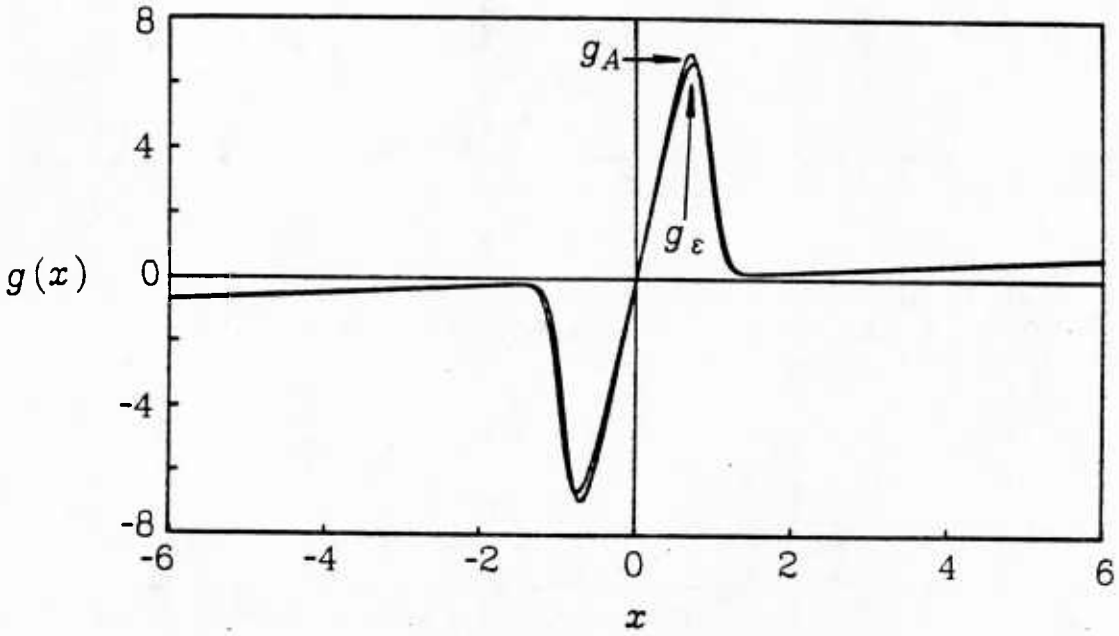

Fig. 2.14. Locally optimal nonlinearity g_A for Middleton Class A density with $A = .1111$ and $\Gamma = .0909$ compared to locally optimal nonlinearity g_ε for Gaussian-Gaussian ε-mixture density with $\sigma_1^2/\sigma_0^2 = 100$ and $\varepsilon = .10$. These values satisfy Eqns. (2.19) and (2.20).

different contaminants of a nominal Gaussian background, including Laplace contamination [23]. This mixture density also figures importantly in Huber's theory of robustness [34], where the contaminant density is merely assumed to have log-convex shape.

Laplace Density

This density is also known as the double sided exponential density, and may be written as

$$f_L(x) = \frac{\sqrt{2}}{2} e^{-\sqrt{2}|x|} \tag{2.20}$$

The LO detector associated with the Laplace density is

$$g_L(x) = \text{sgn}(x) \tag{2.22}$$

We will refer to (2.22) as the sign detector, *sd*. The Laplace density is a convenient model, for it has simple form. Measurements on ocean acoustic data suggest that the Laplace density may be a good model for certain underwater environments [35,36]. While the density clearly has tails heavier than the Gaussian, it also has a non-Gaussian-like mode. Instead of the smooth, infinitely differentiable mode of the Gaussian density, the Laplace density has a cusp. Figure 2.15 compares the two densities for equal variances.

Generalized Gaussian Density

This family of densities is a generalization of the Gaussian density and may be written as

$$f_c(x) = \frac{\gamma c}{2\Gamma(1/c)} e^{-|\gamma x|^c} \tag{2.23}$$

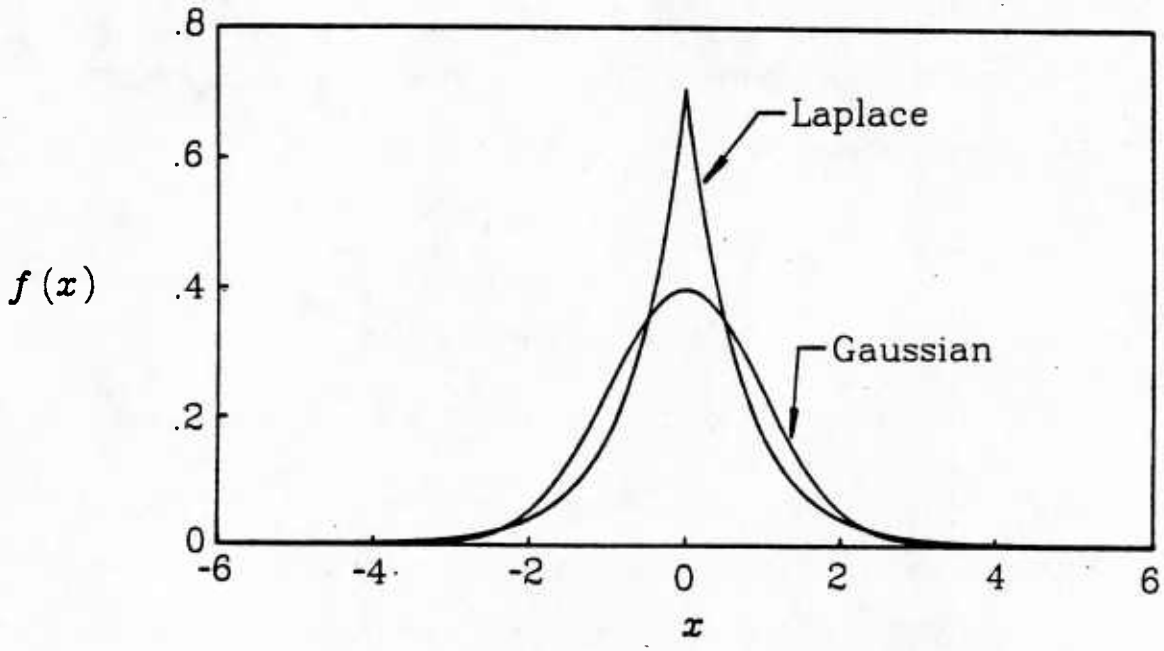

Fig. 2.15. The Laplace density f_L and Gaussian density compared with zero means and unit variances.

where the parameter γ is defined as

$$\gamma = \left[\frac{\Gamma(3/c)}{\Gamma(1/c)}\right]^{1/2}$$

and Γ is the gamma function, given by

$$\Gamma(x) = \int_0^\infty \tau^{x-1} e^{-\tau} d\tau$$

The LO nonlinearity associated with f_c is

$$g_c = c\,\gamma^c\,|x|^{c-1}\,\mathrm{sgn}(x) \qquad (2.24)$$

This family of densities includes the Gaussian density for $c = 2$ and the Laplace density for $c = 1$. It has received attention in previous work, both as a convenient heavy-tailed density for theoretical analyses [17,37-39], as well as a reasonable model for observed noise densities [21]. This family has also been used to describe lighter tailed densities than the Gaussian [36], with values of $c \approx 3$. As $c \to \infty$, the density tends toward a uniform distribution.

Figure 2.16 compares some members of the generalized Gaussian density family on a linear scale, and in Fig. 2.17 they are compared on a logarithmic scale. Some samples of the LO nonlinearity may be found in Fig. 2.18.

Johnson S_u Family – Transformed Gaussian Density

Another family of heavy tailed pdf's is the Johnson S_u family. It has been proposed [40] that certain heavy tailed non-Gaussian densities may be thought of as arising from nonlinear distortions of the Gaussian density. For example, if Y is distributed as a zero mean, unit variance

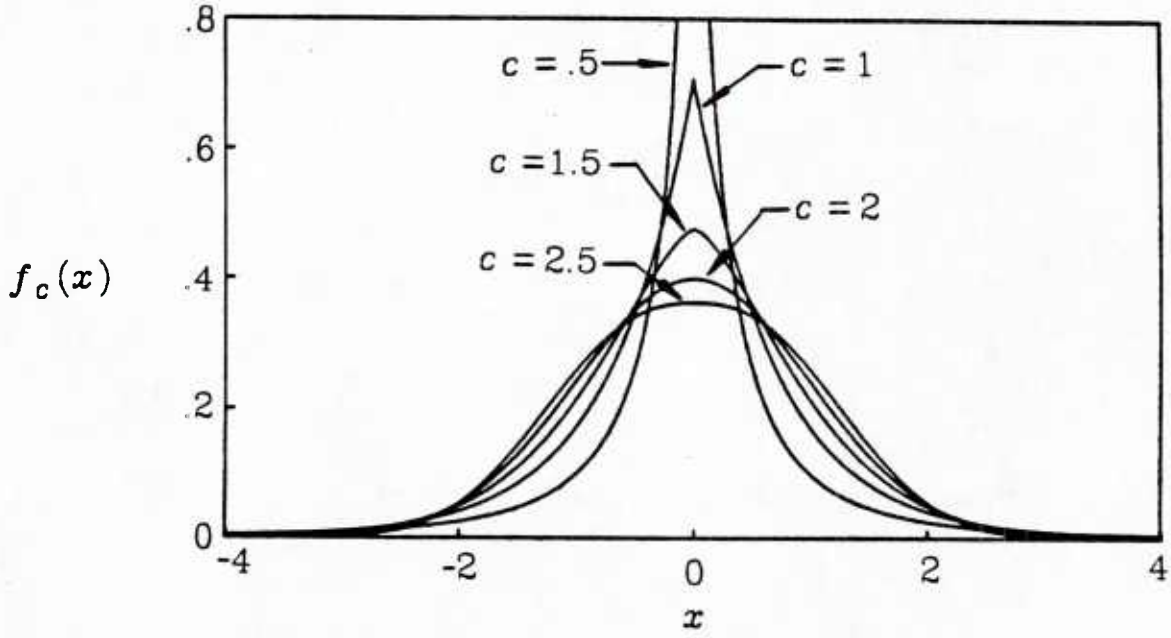

Fig. 2.16. Representative generalized Gaussian densities f_c for various values of parameter c.

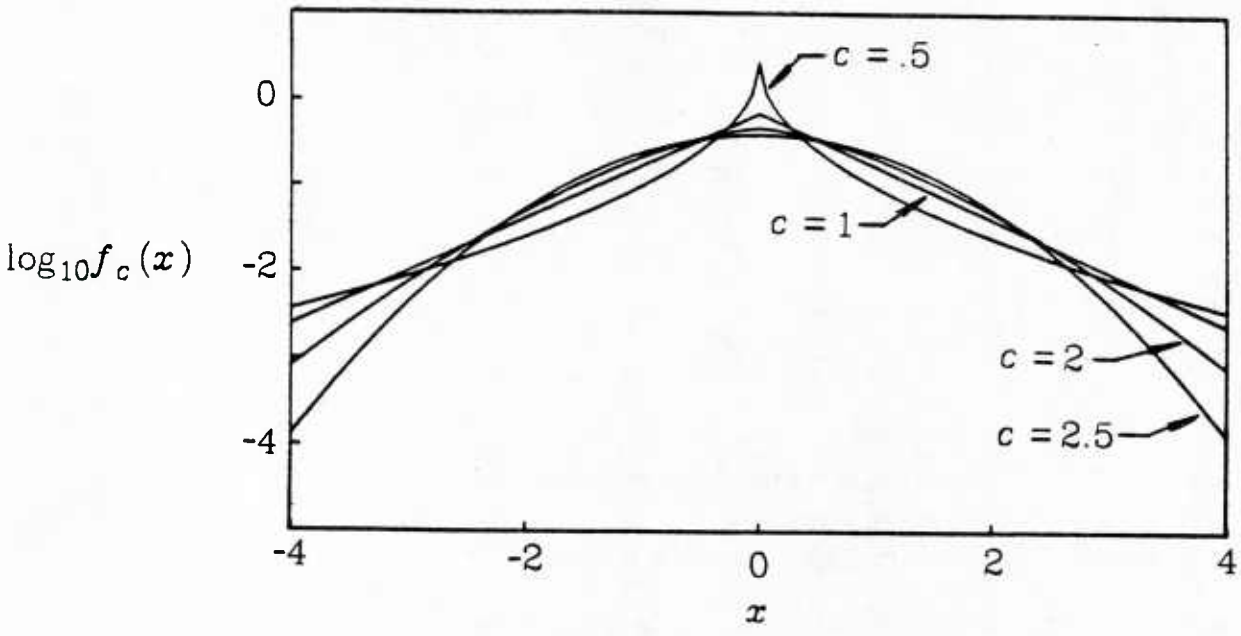

Fig. 2.17. Logarithm of generalized Gaussian densities f_c.

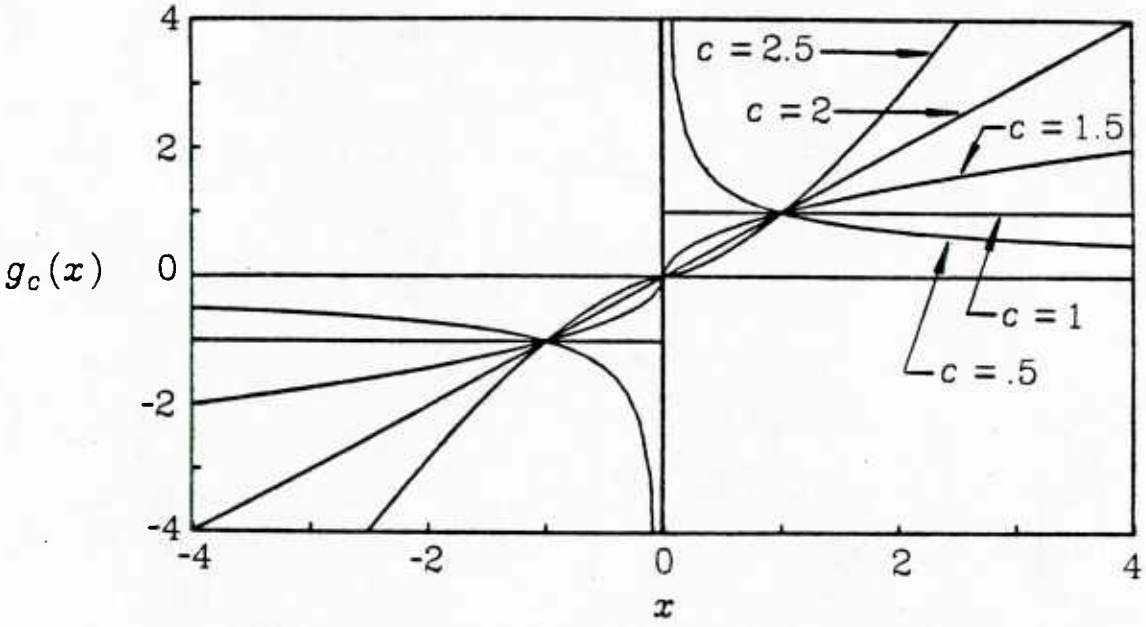

Fig. 2.18. Locally optimum nonlinearity g_c for generalized Gaussian densities.

Gaussian random variable, and a new random variable is defined as

$$X = u \sinh(Y/\delta) \tag{2.25}$$

then the density of X has unit variance, and belongs to the Johnson S_u family, given by

$$f_\delta(x) = \frac{\delta}{u\sqrt{2\pi}}\left[\frac{x^2}{u^2}+1\right]^{-\frac{1}{2}} e^{-(\delta/2)\sinh^{-1}(x/u)^2} \tag{2.26}$$

with

$$u = \left[\frac{2\sigma^2}{e^{(2/\delta^2)}-1}\right]^{\frac{1}{2}} \tag{2.27}$$

The parameter δ controls the tail heaviness. As $\delta \to \infty$, the pdf tails become progressively lighter, and approach Gaussian tails in the limit. Like the generalized Gaussian family, a single parameter indexes the range of tail behaviors.

The LO nonlinearity associated with f_δ may be written as

$$g_\delta(x) = \frac{x}{u^2}\left[1+\frac{x^2}{u^2}\right]^{-1} + \left[1+\frac{x^2}{u^2}\right]^{-\frac{1}{2}}\frac{\delta^2}{u}\sinh^{-1}\frac{x}{u} \tag{2.28}$$

Some representative members of the Johnson S_u family are shown in Fig. 2.19 on a linear scale, and in Fig. 2.20 on a logarithmic scale. The corresponding LO nonlinearities g_δ are given in Figure 2.21.

3. Detectors and the Non-Gaussian Noise Environment

Up to this point, basic detector structures have been reviewed, and some simple noise density models have been presented. We now consider some effects of a non-Gaussian noise environment upon the detection

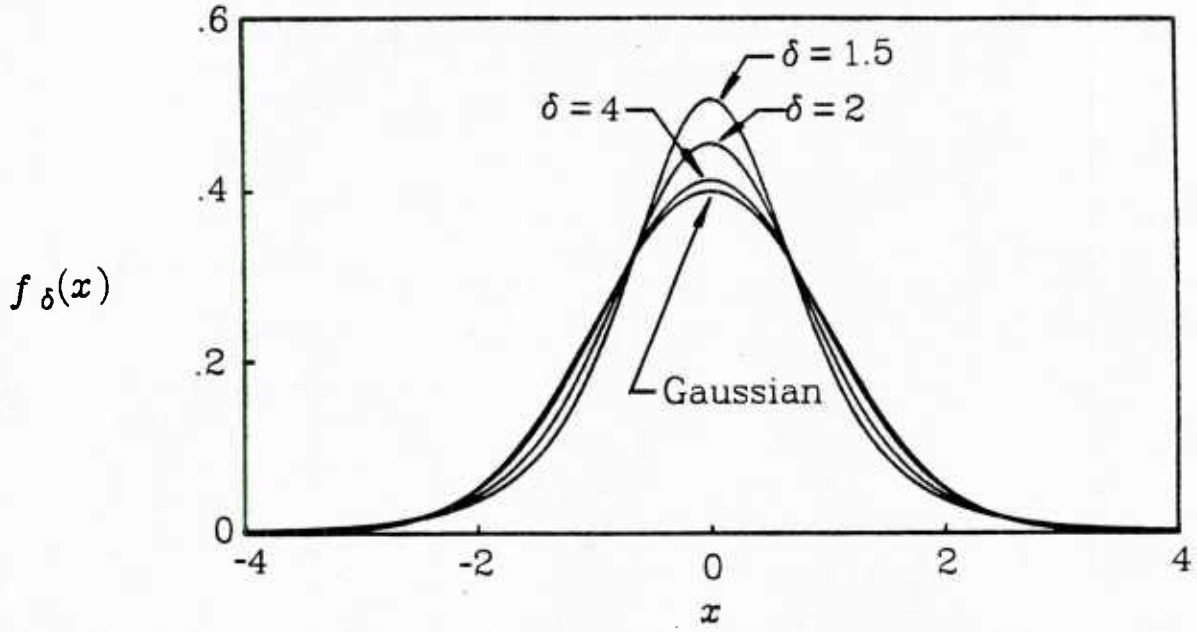

Fig. 2.19. Representative Johnson S_u densities f_δ compared to the Gaussian density.

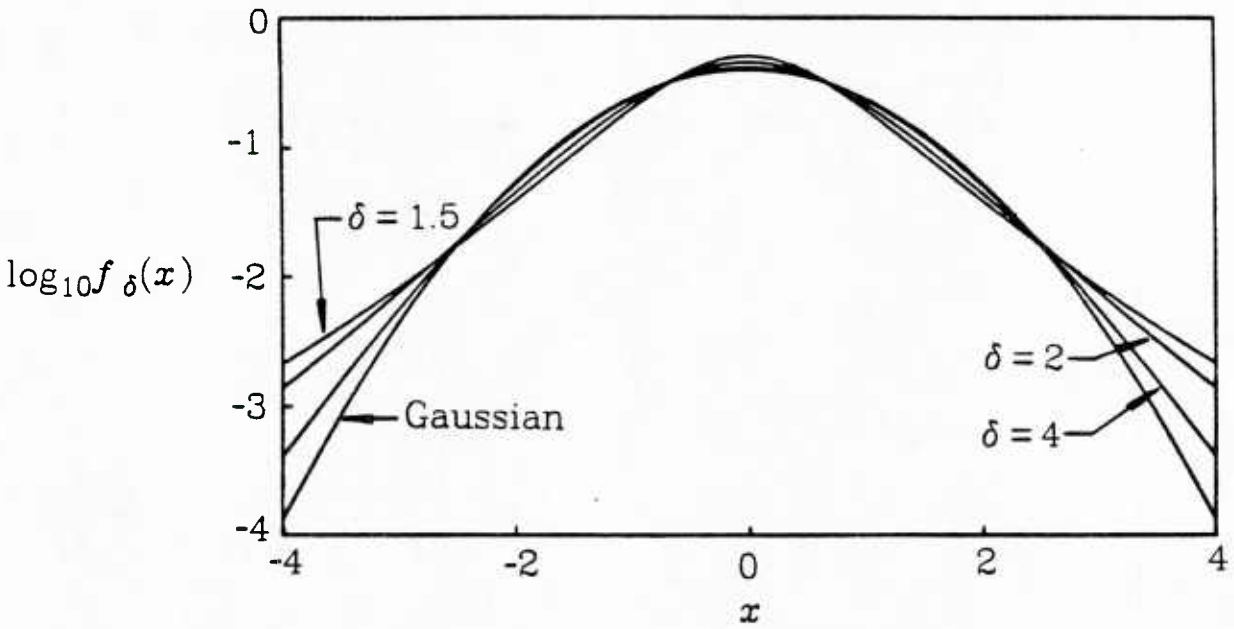

Fig. 2.20 Logarithm of Johnson S_u densities f_δ compared to log of Gaussian density.

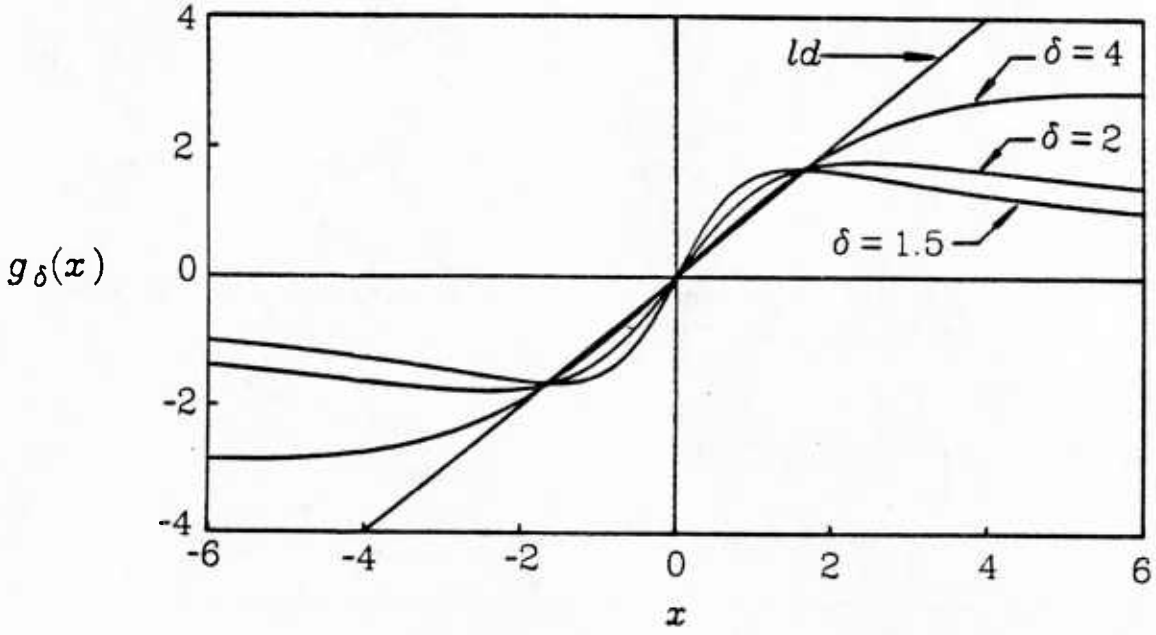

Fig. 2.21. Locally optimum nonlinearity g_δ for Johsnson S_u densities compared to unity slope linear detector ld.

problem.

If two noise processes with equal variances are compared, one Gaussian and the other non-Gaussian in the sense previously discussed, it will become apparent that the non-Gaussian noise has many more large valued observations, or a larger degree of scatter. In the estimation or regression contexts, one might say that the non-Gaussian noise process observations contain a larger number of *outliers*.

Relation between Non-Gaussian Estimation and Detection

Work in robust estimation has long suggested that, in heavy tailed noise, a robust estimator of the mean should reduce the *influence* of very large data observations while leaving observations near the mean relatively unchanged [41]. Any estimator uses a finite number of observations, and an excessive number of outliers unduly affects the estimate, generally increasing its variance with respect to a robust estimator. Note that "excessive", as used here, is a qualitative term, with a meaning dependent upon the particular estimator. Estimators based upon Gaussian noise statistics typically have little protection against outliers, for the simple reason that the effect of large observations is undiminished in any way; even the addition of a single observation with very large magnitude relative to the rest of the observations may significantly distort the outcome.

The effect of an outlier on an estimator can be measured through the calculation of a *sensitivity curve*. Andrews, *et al.* [31], present numerous examples of the sensitivity curves for some common estimators. In estimation of the mean, it turns out that the optimal estimator

has a sensitivity curve given by $g_{OE}(x) = -\frac{\partial}{\partial x}\ln f(x)$. This expression is identical to (2.7), the formula for g_{LO}. Considering the duality between estimation and detection, this is not a surprising result. Further, the Cramer-Rao inequality [1 p. 127] evaluates the efficiency of an estimator g via an expression identical to the efficacy of detector nonlinearity g.

Since the test statistic in a detector also uses only a finite number of observations, the detector nonlinearity must reduce the impact of outliers. NP and LO detector nonlinearities related to non-Gaussian heavy tailed densities are typically composed of a linear region surrounded by tails which compress, limit, or even blank large data observations. The previous examples of LO nonlinearities for some common heavy-tailed noise models exhibit this type of behavior.

Non-Gaussian Density Characterization

Given previously, and repeated here, is the loose definition of the non-Gaussian noise densities which will be of interest in the following chapters: the noise pdf is unimodal, symmetric about its mean placed at the origin, and has nonzero support over the entire real line. Near the mean, the density has a Gaussian-like shape, and the tails asymptotically decrease to zero, but at a slower rate than the Gaussian; *i.e.*, $\lim_{|x|\to\infty} e^{x^2/2\sigma^2} f(x) = \infty$ where σ^2 is the noise variance. Note that both Middleton's Class A density and the Gaussian-Gaussian ε-mixture satisfy this definition, despite being the sums of various Gaussian densities.

The following characteristics loosely specify the LO detector nonlinearities related to the desired types of non-Gaussian densities:

(*a*) continuous, with continuous low-order derivatives

(*b*) approximately linear at the origin

(*c*) odd symmetric about the origin

(*d*) strictly positive to the right of the origin

(*e*) monotone in the tail regions

Note that, in light of (2.7), specification of the LO nonlinearity behavior is equivalent to specifying the form of the associated density.

Motivation for Nearly Optimal Detection

Implicit in both the NP and LO detection methods is a requirement that the noise pdf must be known exactly. This knowledge is needed in order to construct g_{NP} or g_{LO}. In general, the noise statistics are not known with precision and the design of the LO or NP detector is not straightforward. An additional consideration is that often the noise environment is nonstationary, and an adaptive structure is necessary.

Alternative detection strategies are available, and among these are (1) detectors which are robust with respect to deviations from a nominal noise environment [34,37,42,43]; (2) nonparametric detectors which use only very general information about the underlying noise distribution [44,45]; and (3) fixed suboptimal detectors with acceptable performance [27,46-48]. There are some problems with each of the three strategies: first, it is not clear in the design of minimax robust detectors what density should be chosen as the nominal environment and what class of densities should be chosen as unfavorable alternatives. Also, solution of the problem may be quite difficult. Second, while nonparametric methods

are usually simple and afford some degree of protection against heavy tailed noises, they may not be as efficient as possible. Third, a fixed suboptimal detector may be simple to implement, but may suffer severe performance degradation should the noise environment change from nominal conditions.

With these ideas about the non-Gaussian noise environment in mind, the following chapters explore methods related to nearly optimal, yet simple, detector nonlinearities. The previous discussion should make clear the necessity for simple methods which can adapt the detector structure to unknown, and possibly changing, noise environments.

Appendix 2.1

Throughout the thesis, reference will be made to some *Arctic under-ice noise data*. This data is the digitized output of a hydrophone suspended beneath an ocean ice surface. Details of the data collection and analysis of the data is provided by Dwyer [22].

The data, covering a time span of approximately 10 minutes, consists of 6006 records of 1024 data points, sampled at a 10 kHz rate. As Dwyer points out, the data taken as a whole appears to be nonstationary and non-Gaussian: upon further examination, however, it appears that only certain of the noise records deviate significantly from a nominal Gaussian distribution. The following argument is raised:

Define the estimated mean of data record k as

$$\mu_1(k) = \frac{1}{1024} \sum_{i=1}^{1024} n_{k,i}$$

where $n_{k,i}$ is the i^{th} sample of record k. The r^{th} central moment of data record k may be computed for $r > 1$ as

$$\mu_r(k) = \frac{1}{1024} \sum_{i=1}^{1024} [n_{k,i} - \mu_1(k)]^r$$

Using the second, third and fourth central moments, the *skewness* β_1 and the *kurtosis* β_2 of a sample distribution may be computed as

$$\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

Then for each record k, the sample mean μ_1, sample variance σ^2, sample

skewness β_1, and sample kurtosis β_2 may be plotted as a function of k. Examination of the plots, and of the kurtosis plot in particular, reveals that a proportion of the sample records deviate from nominal values. The nominal kurtosis is approximately 3, the exact value of kurtosis for the Gaussian density. Occasionally, values $\beta_2 \gg 3$ are observed. It is these records which are of interest in the thesis, for a high kurtosis value for a unimodal density indicates a heavy-tailed density. For example, the kurtosis of the Laplace density is 6. The sample cdf of the data indicates that the density is unimodal; therefore the conclusion is that the records with a high sample kurtosis have a heavy tailed non-Gaussian density.

The data from records with kurtosis exceeding 4 was collected for use in simulation in the following chapters. Of the 6006 records, 58 met the selection criterion. Figure A2.1 presents a list, indexed by BLOCK, of their record numbers (RECORD), sample means (MEAN), variances (VARIANCE), skewness (BETA1) and kurtosis (BETA2). Figs. A2.2 - A2.5 present this data in graphical form. Data samples from a "typical" high kurtosis block is presented in Fig. A2.6. Note that the deviation from a nominal Gaussian density is apparently confined to two regions where the data has a much greater spread than the majority of data in the block.

BLOCK	RECORD	MEAN	VARIANCE	BETA1	BETA2
1	41	0.1680	0.2243	-0.6789	17.6774
2	48	0.1553	0.1235	0.0230	5.7160
3	57	0.1629	0.1742	-0.2864	5.5454
4	68	0.1698	0.1955	0.0075	4.2291
5	69	0.1566	0.2046	0.2204	4.2246
6	619	0.1664	0.2026	-0.1819	4.3639
7	708	0.1657	0.1937	-0.2008	4.1753
8	724	0.1649	0.2673	-0.3840	4.1981
9	726	0.1630	0.2570	-0.3672	4.1651
10	730	0.1536	0.3806	-0.8361	4.0926
11	732	0.1641	0.3958	-0.6972	4.1245
12	791	0.1684	0.1871	-0.3765	4.0627
13	867	0.1708	0.2621	-0.4333	4.1756
14	1261	0.1678	0.2210	-0.6956	12.4215
15	1349	0.1622	0.1303	0.2202	4.0059
16	1362	0.1549	0.0918	-3.1228	28.1921
17	1363	0.1694	0.0629	0.1948	5.0415
18	1377	0.1606	0.1097	-0.0032	9.8037
19	1380	0.1703	0.0980	0.0397	4.2458
20	1384	0.1648	0.0968	-0.2784	4.7318
21	1388	0.1692	0.0656	-0.1791	5.6106
22	1474	0.1701	0.0929	0.2068	4.1320
23	1484	0.1664	0.0647	0.2371	4.0226
24	1487	0.1583	0.1700	0.5149	24.1614
25	1544	0.1648	0.1096	0.0005	5.8730
26	1678	0.1676	0.1491	-0.0360	5.6124
27	1888	0.1671	0.1056	0.2052	7.7451
28	1916	0.1695	0.0594	-0.1035	4.3778
29	1918	0.1730	0.0842	-0.0928	4.3102
30	1924	0.1644	0.0675	-0.2626	4.1132
31	1938	0.1654	0.0862	-0.5538	7.1052
32	1943	0.1582	0.0565	0.5076	5.5115
33	1946	0.1698	0.0630	-0.6095	5.1463
34	1962	0.1600	0.0656	0.1034	5.1285
35	2041	0.1596	0.1984	-0.8858	11.2526
36	2042	0.1717	0.2063	0.6156	7.4304
37	2066	0.1709	0.1523	-1.5696	29.5222
38	2107	0.1671	0.1113	0.0973	5.2482
39	2114	0.1636	0.0939	0.3051	4.1472
40	2132	0.1697	0.1412	0.2270	5.0950
41	2177	0.1659	0.1000	-1.0546	13.4157
42	2220	0.1422	0.2481	-2.7230	22.9806
43	2226	0.1672	0.1857	-0.2045	7.0628
44	2229	0.1603	0.1425	0.1391	12.5198
45	2230	0.1737	0.1489	0.2172	6.1677
46	2233	0.1648	0.1043	0.1070	4.1405
47	2236	0.1592	0.1963	-1.3927	14.5487
48	2238	0.1672	0.1460	0.3935	5.7368
49	2239	0.1651	0.1770	0.3023	11.6600
50	2240	0.1640	0.1321	-0.1224	6.1600
51	2242	0.1728	0.1401	0.2159	5.0946
52	2246	0.1645	0.2141	1.0191	10.2487
53	2247	0.1672	0.1828	0.0207	11.8074
54	2248	0.1133	0.3774	-2.9252	18.5053
55	2249	0.1690	0.1384	0.1168	5.0615
56	2250	0.1690	0.1677	0.1859	4.8567
57	2254	0.1642	0.1657	0.1431	8.1808
58	2261	0.1713	0.0788	-0.2939	4.0476

Fig. A2.1. Table of data record sample moments for records with kurtosis exceeding 4. BLOCK is the index of the selected records. RECORD indexes the 6006 data records of 1024 samples.

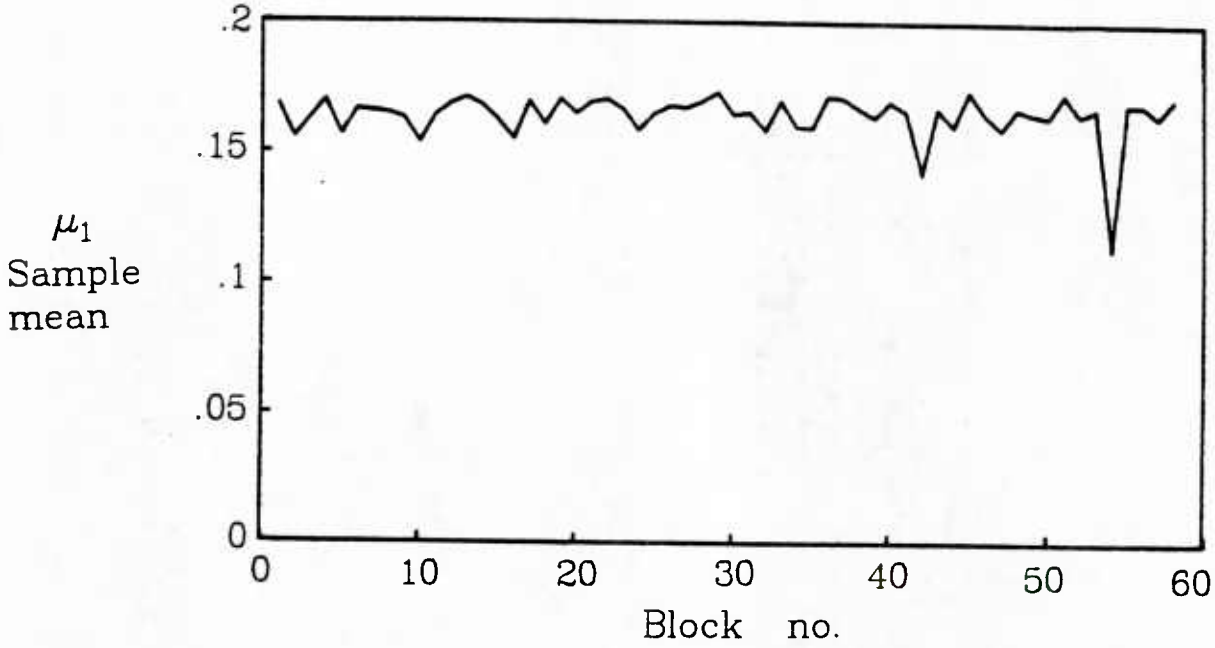

Fig. A2.2. Sample mean μ_1 for the data records with kurtosis exceeding 4. Data records indexed by BLOCK.

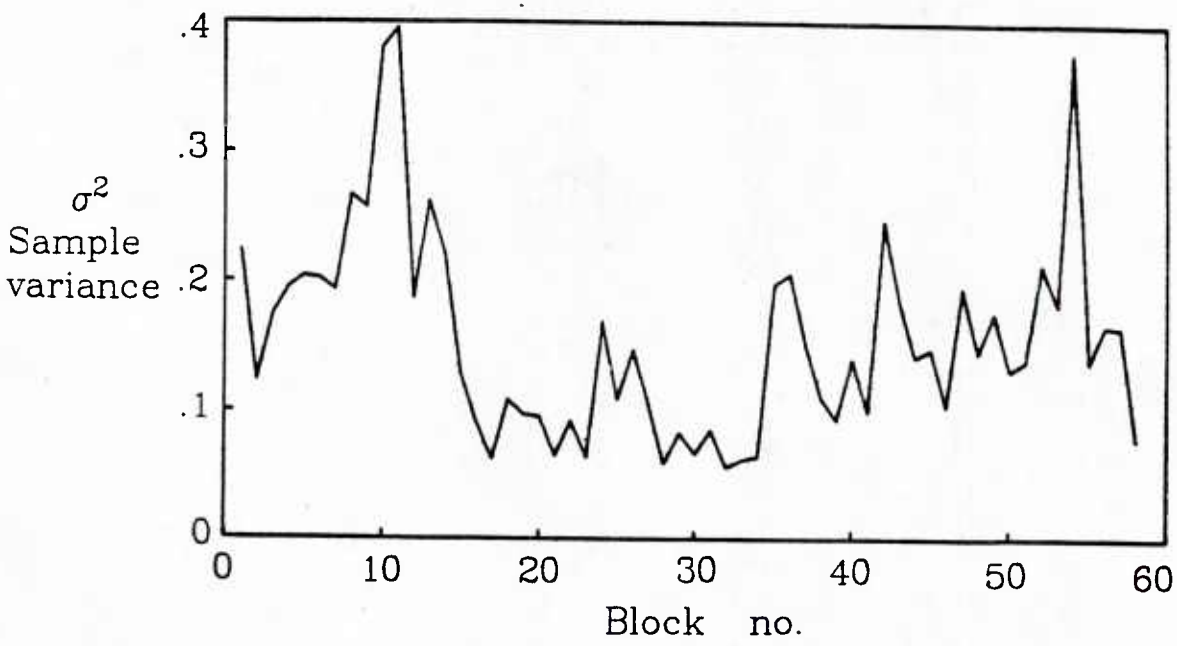

Fig. A2.3. Sample variance σ^2 for the data records with kurtosis exceeding 4. Data records indexed by BLOCK.

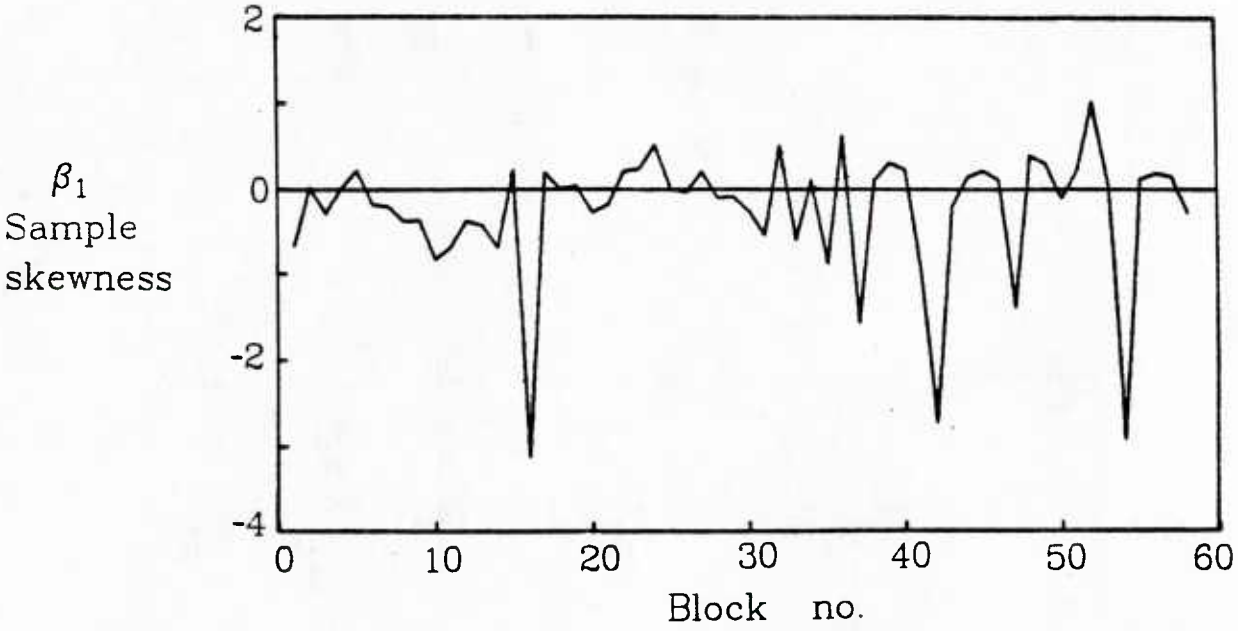

Fig. A2.4 Sample skewness β_1 for the data records with kurtosis exceeding 4. Data records indexed by BLOCK.

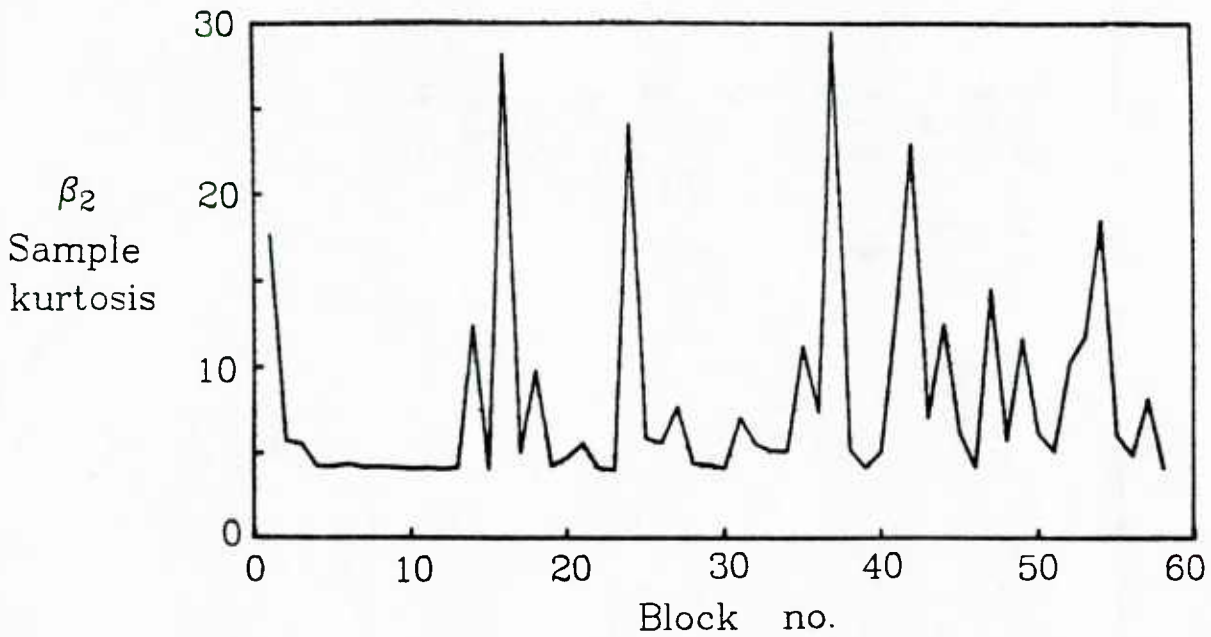

Fig. A2.5. Sample kurtosis β_2 for the data records with kurtosis exceeding 4. Data records indexed by BLOCK.

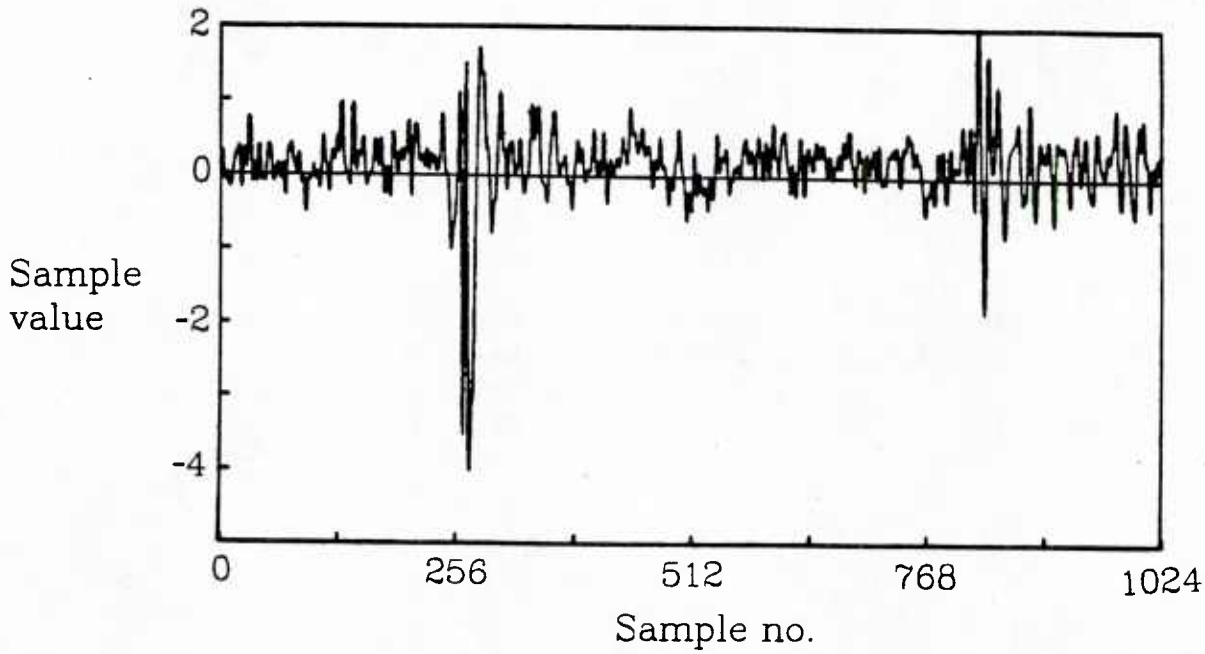

Fig. A2.6. Sample data from data record 2220. The sample moments are $\mu_1 = .1422$; $\sigma^2 = .2481$; $\beta_1 = -2.723$; and $\beta_2 = 22.98$. RECORD=2220, and BLOCK=42.

References

[1] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, Inc: San Francisco, 1977.

[2] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons: New York, 1971.

[3] C.W. Helstrom, *Statistical Theory of Signal Detection*, Oxford; Pergamon Press, 1968.

[4] R.J. Marks, G.L. Wise, D.G. Halverson and J.L. Whited, "Detection in Laplace Noise", *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES- 14, no. 6, pp. 866-871, Nov. 1978.

[5] G.W. Lunk, "Theoretical Aspects of Importance Sampling Applied to False Alarms", *IEEE Trans. Inform Theory*, vol. IT-29, no. 1, pp. 73-82, Jan. 1983.

[6] R.L. Mitchell, "Importance Sampling Applied to Simulation of False Alarm Statistics", *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 1., pp.15-24, Jan. 1981.

[7] R.S. Freedman, "On Gram-Charlier Approximations", *IEEE Trans. Commun.*, vol. COM-29, no. 2., pp. 122-125, Feb. 1981.

[8] C.W. Helstrom, "Approximate Evaluation of Detection Probabilities in Radar and Optical Communications", *IEEE Trans. Aerosp. Electron. Syst.*, AES-14, no. 4, pp. 630-640, July 1978.

[9] D.A. Shnidman, "Efficient Evaluation of Probabilities of Detection and the Generalized *Q*-function", *IEEE Trans. Inform. Theory*, vol. IT-22, no. 6, pp.746-751, Nov. 1976.

[10] H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations", *Annals of Math. Statist.*, vol. 23, pp. 493-507, 1952.

[11] P. Billingsley, *Probability and Measure*, John Wiley and Sons: New York, 1979.

[12] D. Middleton, "Canonically Optimum Threshold Detection", *IEEE Trans. Inform. Theory*, vol. IT-12, no. 2, pp. 230-243, April 1966.

[13] J. Capon, "On the Asymptotic Relative Efficiency of Locally Optimum Detectors", *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67-71, April 1961.

[14] E.J.G. Pitman, *Some Basic Theory for Statistical Inference*, London: Chapman and Hall, 1979.

[15] J.H. Miller and J.B. Thomas, "Numerical Results on the Convergence of Relative Efficiencies", *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-11, no. 2, pp. 204-209, March 1975.

[16] D.L. Michalsky, G.L. Wise, and H.V. Poor, "A Relative Efficiency Study of some Popular Detectors", *J. of Franklin Inst.*, vol. 313, pp. 135-148, March 1982.

[17] J. H. Miller and J. B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 2, pp. 241-250, March 1972.

[18] A. Traganitis, "Narrowband Filtering, Estimation, and Detection of Non-Gaussian Processes", *Ph.D. Dissertation*, Dept. of Electrical Engineering and Computer Science, Princeton University: Princeton, NJ, 1974.

[19] S. L. Bernstein, *et al*, "Long-Range Communications at Extremely Low Frequencies," *Proc. IEEE*, vol. 62, no. 3, pp. 292-312, March 1974.

[20] A. D. Spaulding and D. Middleton, "Optimum Reception in an Impulsive Interference Environment - Part I: Coherent Detection," *IEEE Trans. Commun.*, vol COM-25, no. 9, pp. 924-934, Sept. 1977.

[21] A. D. Watt and E. L. Maxwell, "Measured Statistical Characteristics of VLF Atmospheric Noise," *Proc. Inst. Radio Engineers*, vol. 45, no. 1, pp 55-62, Jan. 1957.

[22] R. F. Dwyer, "Arctic Ambient Noise Statistical Measurement Results and Their Implications to Sonar Performance Improvements," Technical Report 6739, Naval Underwater Systems Center, New London, CT, May 5, 1982.

[23] J.H. Miller and J.B. Thomas, "The Detection of Signals in Impulsive Noise Modeled as a Mixture Process", *IEEE Trans. Commun.*, vol COM-24; no. 5, pp. 559-563, May 1976.

[24] M. Csörgö and P. Révész, *Strong Approximations in Probability and Statistics*, Academic Press: New York, 1974.

[25] D. Middleton, "Statistical-Physical Models of Electromagnetic Interference", *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106-127, August 1977.

[26] D. Middleton, "Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance", *IEEE Trans. Electromagn. Compat.*, vol. EMC-21, no. 3, pp. 209-220, August 1979.

[27] K.S. Vastola, "Threshold Detection in Narrowband Non-Gaussian Noise", Tech. Report no. 46, Dept. of EECS, Princeton University, Princeton NJ, 1983. (To appear in *IEEE Trans. Commun.*)

[28] L.A. Berry, "Understanding Middleton's Canonical Formula for Class A Noise", *IEEE Trans. Electromagn. Compat.*, vol. EMC-23, no. 4, pp. 337-343, Nov. 1981.

[29] M. Abramowitz and I. Stegun, Ed., *Handbook of Mathematical Functions*, New York, NY: Dover Publications, Inc., 1972.

[30] D. Middleton, "Procedures for Determining the Parameters of the First Order Canonical Model of Class A and Class B Electromagnetic Interference", *IEEE Trans. Electromagn. Compat.*, vol. EMC-21, no. 3, pp. 190-208, August 1979.

[31] D. F. Andrews, *et al*, *Robust Estimates of Location*, Princeton University Press, Princeton, NJ, 1972.

[32] E. J. Modugno III, "The Detection of Signals in Impulsive Noise", *Ph.D. Dissertation*, Dept. of EECS, Princeton University, Princeton NJ, 1982.

[33] K.S. Vastola, "On Narrowband Impulsive Noise", *Proc. 20th Annual Allerton Conference on Communications, Control, and Computing*, pp. 739-748, Oct. 1982.

[34] P. J. Huber, *Robust Statistical Procedures*, Society for Industrial and Applied Mathematics: Philadelphia, PA, 1978.

[35] G.R. Wilson and D.R. Powell, "Experimental and Modeled Density Estimates of Underwater Acoustic Returns", *Applied Research Laboratories Tech. Report ARL-TP-82-38*, University of Texas at Austin, 1982.

[36] F.W. Machell and C.S. Penrod, "Probability Density Functions of Ocean Acoustic Noise Processes", *Applied Research Laboratories Tech. Report ARL-TP-82-37*, University of Texas at Arlington, August 1982.

[37] S.A. Kassam and J.B. Thomas, "Asymptotically Robust Detection of a Known Signal in Contaminated Non-Gaussian Noise", *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp.22-26, Jan. 1976.

[38] M. Kanefsky and J.B. Thomas, "On Polarity Detection Schemes with Non-Gaussian Inputs", *J. Franklin Inst.*, vol. 280, pp. 120-138, August 1965.

[39] J. W. Modestino, "Adaptive Detection of Signals in Impulsive Noise Environments," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 1022-1026, Sept. 1977.

[40] N.L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, John Wiley and Sons: New York, 1971.

[41] F. R. Hampel, "The Influence Curve and its Role in Robust Estimation," *Jour. Amer. Stat. Assoc.*, vol. 69, no. 346, pp. 383-393, June 1974.

[42] A. H. El-Sawy and V. D. Vandelinde, "Robust Detection of Known Signals," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 6, pp. 722-727, Nov. 1977.

[43] R.D. Martin and S.C. Schwartz, "Robust Detection of a Known Signal in Nearly Gaussian Noise", *IEEE Trans. Inform. Theory*, vol. IT-17, no. 1, pp.50-56, Jan. 1971.

[44] J. B. Thomas, "Nonparametric Detection," *Proc. IEEE*, vol.58, no. 5, pp.623-631, May 1970.

[45] S.A. Kassam, "A Bibliography on Nonparametric Detection", *IEEE Trans. Inform. Theory*, vol IT-26, no. 5, pp. 595-602, Sept. 1980.

[46] J. H. Miller and J. B. Thomas, "Robust Detectors for Signals in Non-Gaussian Noise," *IEEE Trans. Commun.*, vol. COM-25, no. 7, pp. 686-690, July 1977.

[47] R.E. Ziemer and R.B. Fluchel, "Selection of Blanking and Limiting Levels for Binary Signaling in Gaussian plus Impulsive Noise", *Proc. IEEE Fall Electronics Conference*, Chicago, IL, pp. 290-295, Oct. 1971.

[48] R.F. Ingram and R. Houle, "Performance of the Optimum and Several Suboptimum Receivers fro Threshold Detection of a Known Signal in Additive, White, Non-Gaussian Noise", *Tech. Report 6339*, Naval Underwater Systems Center, New London, CT, Nov. 24, 1980.

[49] M.H. Meyers, "Computing the Distribution of a Random Variable via Gaussian Quadrature Rules", *Bell System Tech. Journal*, vol. 61, no. 9, pp. 2245-2261, Nov. 1982.

[50] Y.C. Liu and G.L. Wise, "Some Relative Efficiency Results for Detectors for Time Varying Signals", *Proc. IEEE Intl. Conf. on Commun.*, Boston, MA, June 19-22, 1983.

[51] A.B. Martinez, "Asymptotic Performance of Detectors with Non-zero Input SNR", *Proc. 20th Annual Allerton Conf. on Commun., Control, and Computing*, University of Illinois, Urbana-Champaign, IL, pp. 749-758, Sept. 1982.

3

Adaptive Optimization of Suboptimal Nonlinearities

This chapter investigates the feasibility of two simple alternatives to locally optimal detector nonlinearities. Provided some simple measurements on the noise density are available, it is demonstrated that it is possible to construct nonlinearities which produce near-optimal levels of performance in several specific noise environments. Section 1 presents an overview of some practical issues which motivate the necessity for near-optimal, yet uncomplicated, detector nonlinearities.

The basic philosophy forwarded is that nonlinearities designed for practical detectors should have an uncomplicated structure that may be easily adapted to changing noise situations. Two main issues are addressed: the first is development of algorithms to determine the gross shape (input-output relationship) of the nonlinearity. Of primary importance is the tail behavior of the nonlinearity, for it will determine the degree to which impulsively contaminated observations can influence the

detector test statistic. Sections 2 and 3 discuss two alternatives to optimal nonlinearities. The second issue resolved is the matter of scaling the input to the nonlinearity. This problem is essentially equivalent to determining the noise variance and scaling the input. However, the usual estimators of variance depending upon the squares of the noise observations are inefficient when the noise has a heavy tailed density. An alternative scaling method is developed near the ends of Sections 2 and 3.

Sections 4 and 5 provide a numerical comparison of the suboptimal nonlinearities for cases where the true noise density is known. Also, the algorithms are simulated using observed noise data. Section 6 provides a review of the techniques and results presented in this chapter.

1. Introduction

In principle, the design of a Neyman-Pearson (NP) or locally optimal (LO) detector nonlinearity for a signal in additive white noise is a simple matter when the noise statistics are known exactly. There are, unfortunately, some practical problems related to the implementation of a nonlinearity. The most significant problem is simply that the true noise statistics are usually unknown, or changing in time. While well known techniques exist for obtaining the noise density [1-3], they often require a fairly large observation period to achieve an acceptably smooth estimate. For example, Wilson and Powell [4] present kernel function type density estimates of several observed noises. The estimates are noisy and rough looking when the logarithm of the densities are plotted. A LO nonlinearity could be estimated from the derivative of the log of the densities, but this would further emphasize the roughness. Additional smoothing of the

density would be required if an acceptable nonlinearity is to be obtained, and even then the nonlinearity may be still somewhat noisy or rough looking (*e.g.* [5]).

Another problem is that, even when the density is known or can be estimated smoothly, the related memoryless nonlinearity (ZNL) itself is sometimes complicated enough to make implementation or adaptation relatively difficult. For example, the Middleton Class A and Class B noise models have been proposed as physically based canonical representations for non-Gaussian noise, with parameters that may be calculated directly from physical considerations. Both models are infinite series [6,7]; the Class A series comprises weighted Gaussian density terms, while the Class B series comprises confluent hypergeometric functions, which themselves are defined generally via an infinite series [8, p.504]. The detector nonlinearities associated with these models may be calculated directly, but at the expense of a high computational burden. Adaptation of the nonlinearities incurs a similar computational cost.

One approach toward overcoming these difficulties with the optimal nonlinearity is to use a suboptimal ZNL that has nearly optimal performance, but has a structure that is simple to implement and easily adaptable. Some recent examples of this approach include the work by Miller and Thomas [9], Modestino [10], Ingram and Houle [11], Ziemer and Fluchel [12], and Vastola [18,19].

This chapter presents an approach to the design of a noise-adaptive suboptimal detectors with these ideas in mind, focusing on the locally optimal detection problem and noise environment of Chapter 2.

2. Approximation via Noise Tail Matching

The previous chapter presented a discussion of a particular type of non-Gaussian noise environment where the mode of the noise pdf appeared as Gaussian-like, but the density tails were much heavier than the Gaussian. As noted, the LO nonlinearities associated with these types of densities have a nearly linear processing characteristic for input values near the pdf mode. On the other hand, the tail behavior of the LO nonlinearities ranges from linear for a noise pdf with Gaussian tails, to a limiter for exponentially decreasing pdf tails, to a blanker for algebraically decreasing pdf tails. In general, the heavier-tailed the noise density is relative to the Gaussian pdf, the more severely curtailed is the effect of large data observations.

One objective of a noise adaptive nonlinearity, then, should be to relate the ZNL tail behavior to the actually observed noise pdf tail behavior. The main idea of the algorithm in this section is to establish a relation between a measure of tail heaviness and a member of a convenient class of heavy tailed densities. Rather than performing a parametric fitting within the density class, the algorithm chooses a density whose tail characteristics have the same tail heaviness measure as the actual noise density. The nonlinearity tails are thus determined by the member of the density class chosen. The central region of the nonlinearity joins the two tails with some function which gives a desirable near-linear processing.

Tail Selection Procedure

This idea is clarified and illustrated by proposing the following: It has been reported by Watt and Maxwell [21] that the generalized Gaussian density

$$f_c(x) = \frac{\gamma c}{2\Gamma(1/c)} e^{-|\gamma x|^c} \tag{3.1}$$

in certain instances can describe the pdf tails of physical noise sources. For a noise variance of σ^2, the parameter γ is defined by

$$\gamma = \left[\frac{\Gamma(3/c)}{\sigma^2\Gamma(1/c)}\right]^{1/2}$$

The corresponding LO nonlinearities, shown earlier in Fig. 2.18, may be written as

$$g_{LO}(x) = c\gamma^c |x|^{c-1} \mathrm{sgn}(x) \tag{3.2}$$

with c conveniently parameterizing ZNL tail behavior. Therefore, we model the observed noise pdf tails via the generalized Gaussian family. If these density tails are used to generate a suboptimal LO nonlinearity, it will have power law tails described by

$$g_{tm}(x) = \hat{c}\gamma^{\hat{c}} |x|^{\hat{c}-1} \mathrm{sgn}(x) \quad \text{for} |x| > x_0 \tag{3.3}$$

It is necessary to find a value $\hat{c}$ such that $f_{\hat{c}}$ is a good approximation to the tail behavior of the true, but unknown, underlying noise density. A simple way to do this is to equate the tail probability of $f_{\hat{c}}$ with the observed tail mass

$$\hat{P}_T = \frac{1}{N}\sum_{i=1}^{N} I_{(T,\infty)}(|r_i|) \tag{3.4}$$

Here, I is the indicator function and r_i are the noise observations presumed available from a noise reference channel. The exponent $\hat{c}$ may be estimated as the value giving

$$2\int_T^\infty f_{\hat{c}}(x)dx = \hat{P}_T \tag{3.5}$$

where, for convenience, it is assumed that the noise has zero mean and unity variance. The estimate $\hat{c}$ is defined implicitly by the integral in (3.5); therefore it is desirable to derive a simpler explicit relationship

$$\hat{c} = h_T(\hat{P}_T) \tag{3.6}$$

One obvious method for obtaining (3.6) is to first calculate P_T as a function of c, and then use interpolation to find the inverse relation h_T. This tabulated version of h_T is shown in Fig. 3.1.

With σ^2 fixed and c small, the value of γ, a scale factor, becomes large. Even though $f_c(x)$ approaches zero asymptotically at a much slower rate than the Gaussian pdf as $|x|$ becomes large, the total probability mass in the tails is quite small. As a result, h_T is multiple valued, the density is extremely peaked, and the LO nonlinearity has a discontinuity at the origin. For $c < 1$, the requirement that the suboptimal ZNL be nearly linear at the origin clearly is not met by (3.2).

The objective in using the generalized Gaussian pdf is to relate the tail heaviness of an observed noise to a parameter governing the shape of the ZNL tail. Therefore we replace the anomalous behavior of the true function h with a simple linear relation

$$\hat{h}_T(\hat{P}_T) = k_1\hat{P}_T + k_2 \tag{3.7}$$

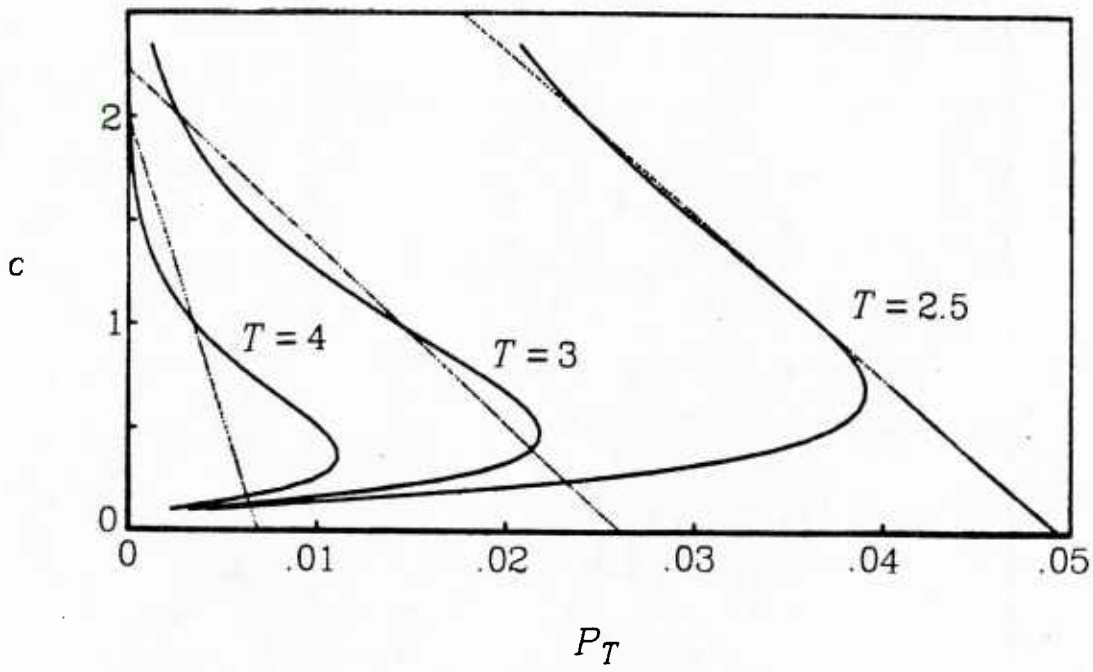

Fig. 3.1. The exact and approximate relationships between exponent c and tail probability P_T for unit variance generalized gaussian density for various thresholds T. The exact relations h_T are the solid curved lines, and the linear approximations $\hat{h}_T$ are the broken straight lines.

where k_1 and k_2 are chosen to approximate (3.6) for a particular value of T. Several sample approximations are plotted as the broken lines on Fig. 3.1. The values for k_1 and k_2 are chosen so that when $\hat{P}_T$ corresponds to Gaussian or exponential noise tails, (3.7) gives $\hat{c} = 2$ and $\hat{c} = 1$, respectively. Note that the linear relation allows the value of $\hat{c}$ to be negative for large tail probabilities.

The tail measurement threshold T must be chosen prior to estimating parameters k_1 and k_2. One way to pick T is to choose a value solving

$$\min_T \mathrm{E}_c\{\mathrm{var}\,\hat{c}\} = \min_T \frac{a^2}{N}\mathrm{E}_c\{P_T(1-P_T)\} \tag{3.8}$$

for some prior density on the parameter c, where N is the number of noise observations. For c uniformly distributed on the interval [1,2], the value $T = 3\sigma$ approximately minimizes (3.8). In practice, some better knowledge of the distribution of c should develop, and T may be adjusted to minimize (3.8).

Central Region Selection

The LO nonlinearities of the generalized Gaussian family have desirable tail behavior, but for small values of $\hat{c}$ the behavior does not meet the constraint of linearity near the origin. To eliminate this behavior, the ZNL needs modification in the region near the origin. A way in which to do this is to replace $g_{tm}(x)$, for x near zero, with a function that will smoothly connect the two tails and have linear-like behavior near the origin. A suitable family of functions are polynomials $p(x)$ with the following characteristics:

$$
\begin{aligned}
&p(x) = \alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_0 \quad \text{for} \quad 0 \le x \le x_0 \\
&p(0) = 0 \\
&p(|\pm x_0|)\mathrm{sgn}(\pm x_0) = g_{tm}(\pm x_0) \\
&p'(|\pm x_0|) = g'_{tm}(\pm x_0) \\
&p''(|\pm x_0|)\mathrm{sgn}(\pm x_0) = g''_{tm}(\pm x_0)
\end{aligned}
$$

Also, because $p(x)$ is a third order polynomial, $p'(x) \approx \alpha_1$ for $|x|$ very near zero. This implies that p will be nearly linear in a neighborhood about the origin, for there its slope is approximately independent of x.

Scaling

The choice of tail behavior via $\hat{c}$ and the point x_0 completely specify $p(x)$. The method for choosing $\hat{c}$ has already been specified, leaving x_0 as the sole free parameter. A method equivalent to choosing the proper x_0 is to choose an arbitrary x_0 and scale the input to the ZNL with a factor ν. It is reasonable to choose ν to maximize the efficacy of the ZNL. For an arbitrary nonlinearity g, efficacy as a function of ν may be rewritten as

$$
\eta_f(\nu) = \frac{\nu^2 \mathrm{E}_f^2[g'(\nu x)]}{\mathrm{E}_f[g^2(\nu x)] - \mathrm{E}_f^2[g(\nu x)]} \tag{3.9}
$$

In principle, (3.9) can be solved exactly. Unfortunately, a closed form solution for ν cannot be found in general, and the density f is generally unknown. These problems may be circumvented by approximating the expectations with integrations over the noise empirical distribution, and solving (3.9) via stochastic approximation methods.

At this point, specification of the suboptimal nonlinearity g_{tm} is complete, and may be written as

$$
g_{tm}(\nu x) = \begin{cases} p(|\nu x|)\,\mathrm{sgn}(\nu x) & \text{if } |\nu x| \le x_0 \\ \hat{c}\,|\nu x|^{\hat{c}-1}\mathrm{sgn}(\nu x) & \text{if } |\nu x| > x_0 \end{cases} \tag{3.10}
$$

Figures 3.2 to 3.4 give some examples of the types of nonlinearities available using this approximation method.

3. Optimization via Efficacy Maximization

The previous section developed a method for choosing suboptimal nonlinearities in what is primarily an "open-loop" fashion: A relation was established *a priori* between observed tail heaviness and tail heaviness of known noise densities. The tails of the known densities were then used to generate tails for the suboptimal ZNL. It is not obvious that this method is optimal in any sense, save for its sheer simplicity.

Another approach is to choose a class of nonlinearities of desirable shape and convenient parameterization, and then find the member of the class which maximizes performance. This type of approach may be considered to be a "closed-loop" technique, for measurements on the observed noise density lead to selection of the optimal member of the nonlinearity class; the performance measure provides "feedback" to the selection algorithm.

Again, under the detection situation and noise environment described in Chapter 2, the following suboptimal ZNL is proposed, comprising a central linear region and two linear tail regions:

$$g_{2l}(x) = (\operatorname{sgn} x)\begin{cases} |x| & \text{for } |x| < a \\ b|x| + a(1-b) & \text{for } a < |x| < x_T \end{cases} \tag{3.11}$$

where

$$x_T = \begin{cases} \dfrac{a(b-1)}{b} & \text{for } b < 0 \\ \infty & \text{for } b \geq 0 \end{cases} \tag{3.12}$$

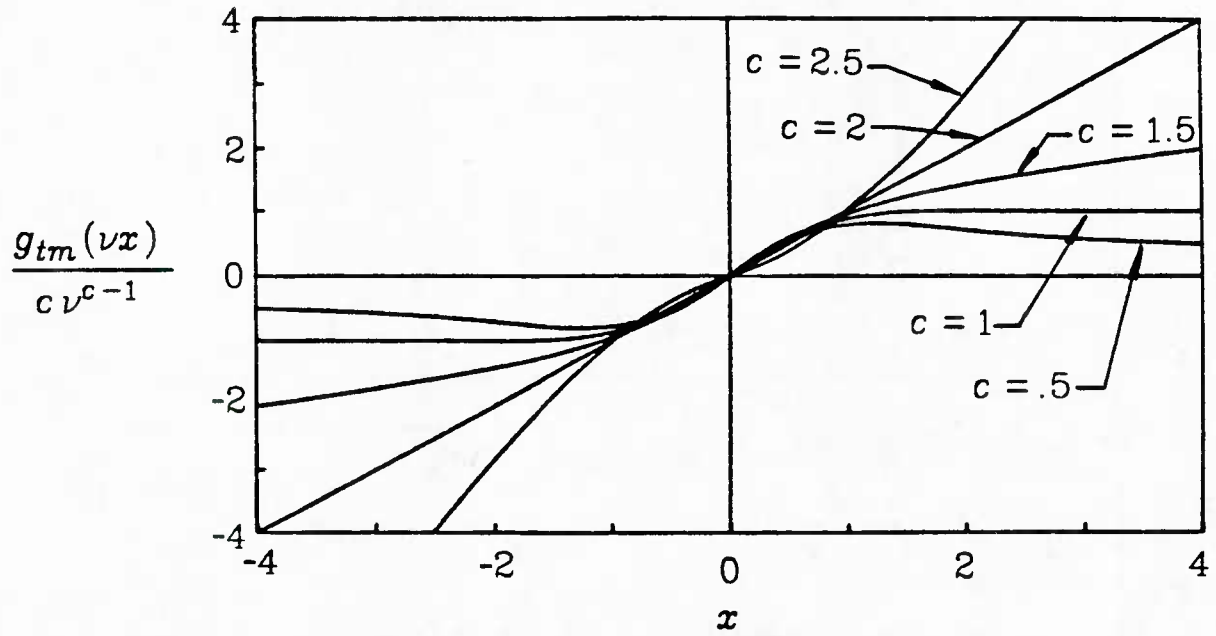

Fig. 3.2. Representative nonlinearities g_{tm} for $x_0 = 3$, $\nu = 1$, and various c.

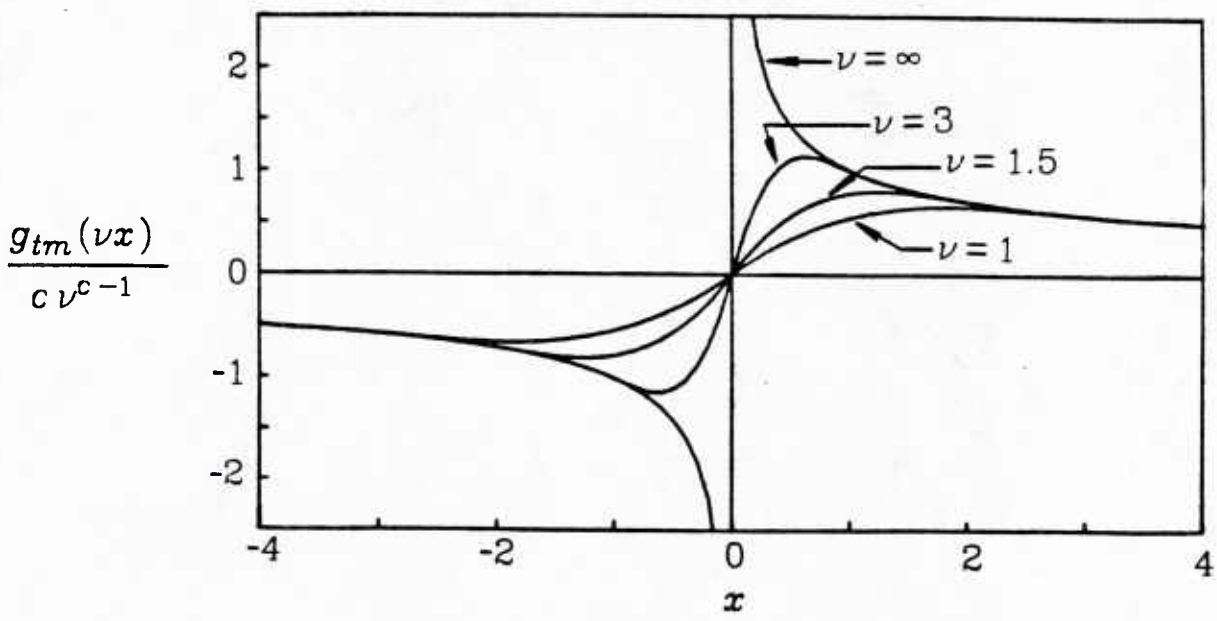

Fig. 3.3. Representative nonlinearities g_{tm} for $x_0 = 3$, various ν, and $c = .5$.

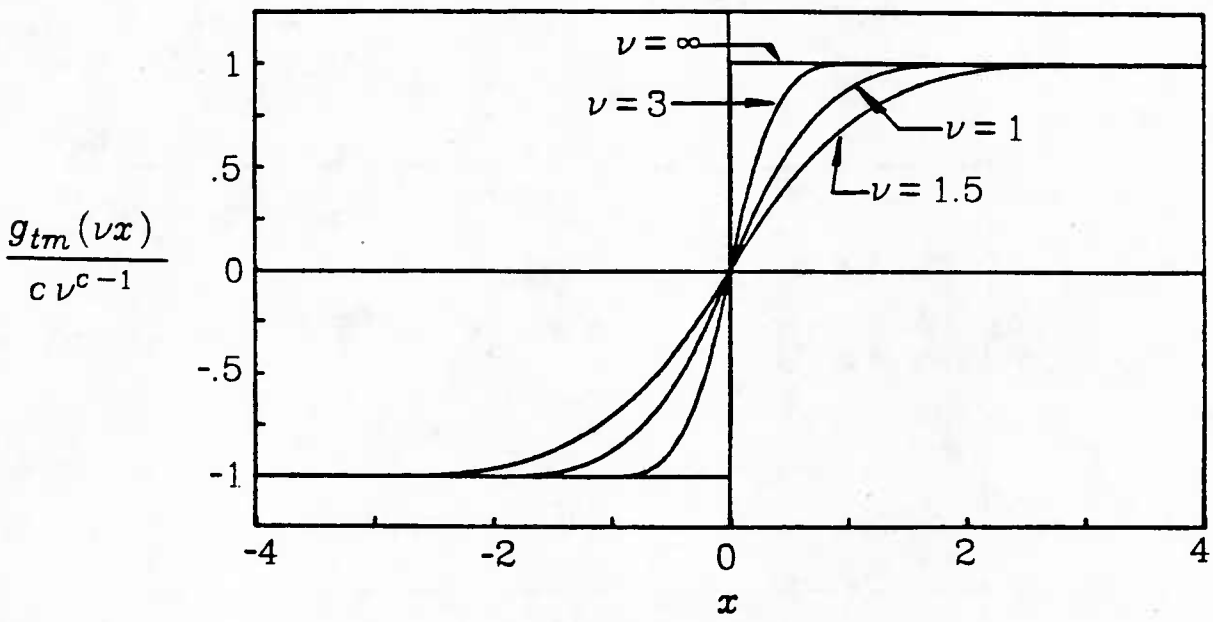

Fig. 3.4. Representative nonlinearities g_{tm} for $x_0 = 3$, various ν, and $c = 1$.

The parameter a governs the breakpoint between the central and tail regions, and b governs the slope of the tail region segments. A pair of representative examples of $g_{2l}(x)$ are given in Figs. 3.5 and 3.6.

Procedure for Estimating Tail Slope

Initially, we will assume a is fixed and turn attention to the problem of estimating b. Some comments will be made in the following subsection on the issue of finding the breakpoint.

The usual performance measure for a LO detector g is efficacy, discussed in Chapter 2 and recalled here as

$$\eta_f(g) = \frac{\left[\int_{-\infty}^{\infty} g'(x) f(x) dx\right]^2}{\int_{-\infty}^{\infty} g^2(x) f(x) dx} \tag{3.13}$$

where the underlying noise density is f and g has zero mean under f.

Figs. 3.5 and 3.6 highlight the fact that there are two distinct possibilities for the shape of g_{2l}: In Fig. 3.5, the slope parameter b is greater than zero, and $g_{2l;b^+}$ is nonzero over the entire tail region. In Fig. 3.6, the parameter b is less than zero, and $g_{2l;b^-}$ is nonzero only over a finite interval.

Using (3.13), the efficacy of $g_{2l;b^+}$ may be written as

$$\eta(g_{2l;b^+}) = \frac{4\left[\int_0^a f + b\int_a^{\infty} f\right]^2}{2\int_0^a x^2 f + 2\int_a^{\infty} (bx + a(1-b))^2 f} \tag{3.14}$$

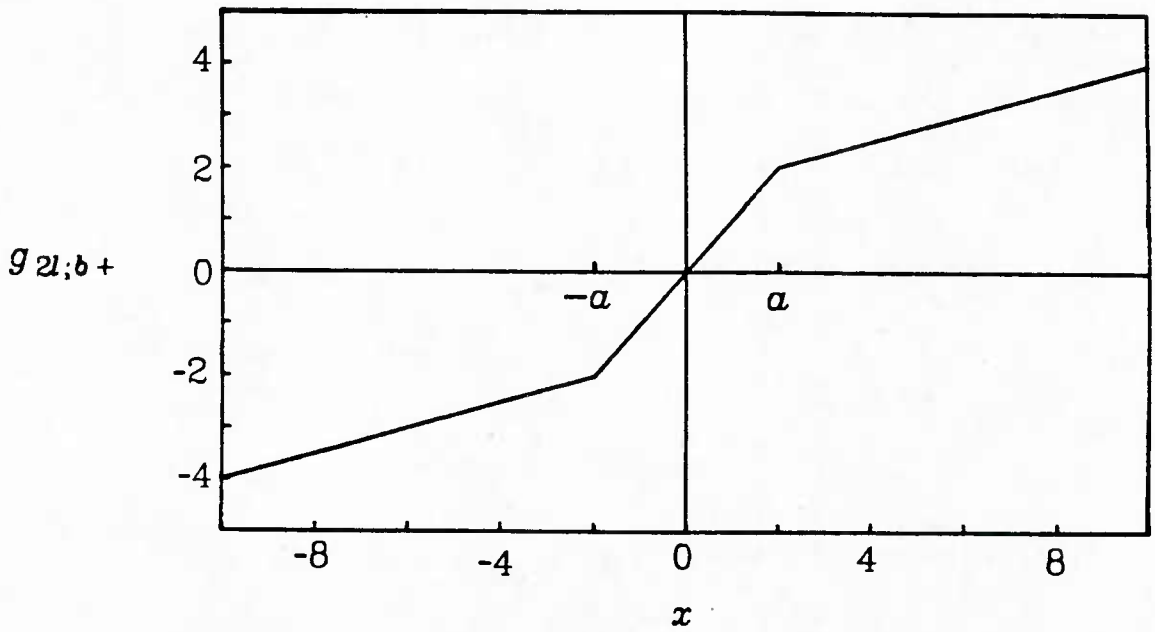

Fig. 3.5. A representative example of $g_{2l;b+}$ for $a = 2$ and $b = .25$.

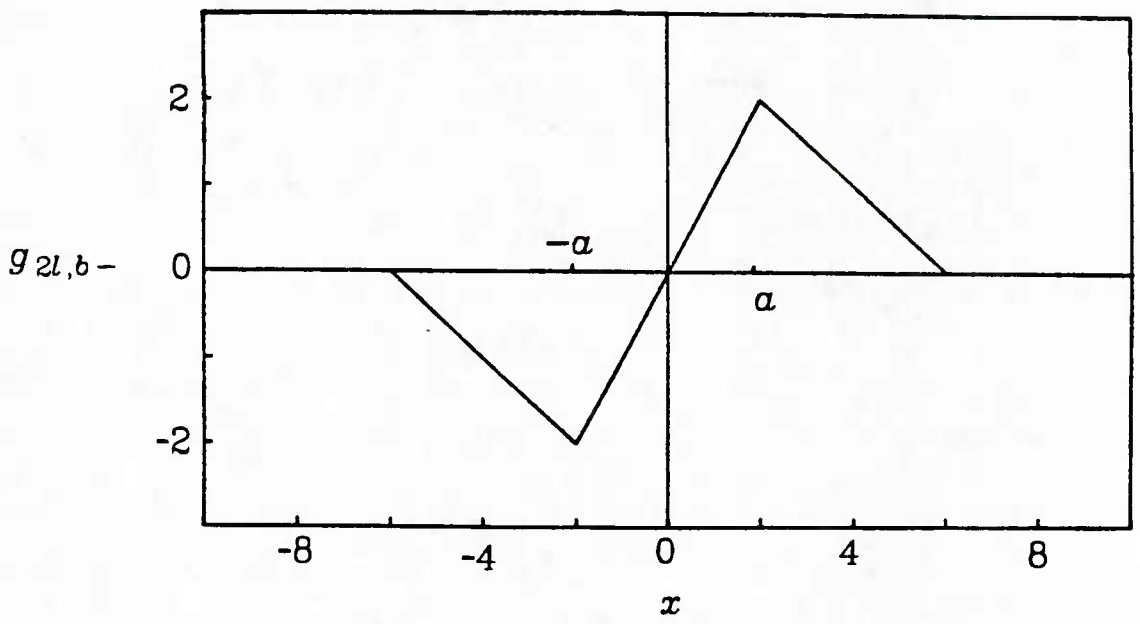

Fig. 3.6. A representative example of $g_{2l;b-}$ for $a = 2$ and $b = -.5$.

Here, $\eta(g_{2l;b^+})$ is an explicit function of b, so it may be maximized by finding b^* such that $\left.\frac{\partial \eta}{\partial b}\right|_{b=b^*} = 0$. In Appendix 3.1, an explicit solution (A3.6) for b^* is found to be a rational function of the partial moments of f, which depend implicitly on the choice of a. It was not possible to find an expression which yields explicitly a value a^* that maximizes (3.14) as a function of a.

The explicit solution (A3.6) was derived to find the b^* that maximizes $\eta(g_{2l;b^+})$. When the solution b^* is non-negative, the tails of the nonlinearity diverge from the x-axis, and g_{2l} has support over the entire real line. Thus, the formulation for efficacy, given in (3.14) is correct, and the solution (A3.6) is correct.

What if (A3.6) yields a result $b^* < 0$? The result b^* is still valid, but the nonlinearity for which efficacy is maximized *is not* $g_{2l;b^-}$. Certain integrals in (3.5) have range of integration (a,∞), whereas the correct expression for the efficacy of g_{2l,b^-} may be written as

$$\eta(g_{2l,b^-}) = \frac{4\left[\int_0^a f + b\int_a^{x_T} f\right]^2}{2\int_0^a x^2 f + 2\int_a^{x_T} (bx + a(1-b))^2 f} \tag{3.15}$$

with x_T given by (3.12). Note that if $b^* < 0$ and the value $x_T = \infty$ is used, then what is actually maximized is the efficacy of a nonlinearity g_V with virtual tails such as those shown in Fig. 3.7.

It is desirable to find an explicit solution for b^* which takes into account the fact that if $b^* < 0$, then the solution should have been

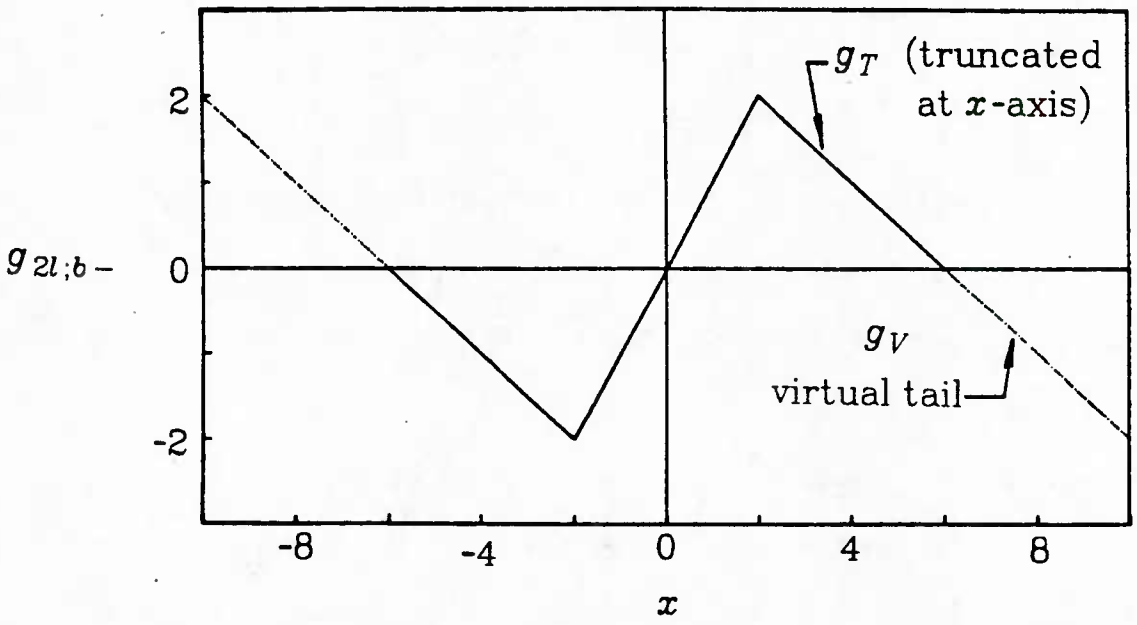

Fig. 3.7. The nonlinearity g_T is the incorrectly optimized $g_{2l;b-}$. It is a truncated version of g_V, whose virtual tails are artifacts of the optimization procedure.

generated by maximizing (3.15) instead of maximizing (3.14). However, the value x_T depends on b^*, and a closed-form expression for b^* could not be developed, as was done for (3.14) where the limits of integration do not depend on b^*.

It is possible to salvage the solution (A3.6). First, note that when (A3.6) is applied without modification and gives $b^* < 0$, the value $|b^*|$ will actually have been underestimated, for this will reduce the mean square error between the virtual tails of g_V and the x-axis, at the expense of increasing the mean square error between the tails dictated by the incorrect b^* and the tails of the properly optimized $g_{2l;b}$- in the interval $[a,x_T]$. Overall, this has the effect of minimizing the performance degradation due to the virtual tails [15]. A further discussion of the mean square error issue in ZNL approximation may be found in Chapter 5.

Two options are available: one is to apply (A3.6), obtain b^*, and if it is less than zero, calculate x_T using (3.12), and merely truncate the nonlinearity at $\pm x_T$. Appendix 3.2 demonstrates that truncating at $\pm x_T$ yields better performance than if the virtual tails were ignored and allowed to remain. The other option is to apply (A3.6) iteratively. For startup, (A3.6) is applied directly, giving b^*_0. Eqn. (3.12) may be used to give an initial value for x_T, and the integrals in (A3.6) may be modified to have range of integration (a,x_T) instead of (a,∞). The appropriately modified (A3.6) gives b^*_1, and the process may continue in this fashion until $|b^*_{n+1} - b^*_n|$ is less than a predetermined accuracy.

Scaling

The issue of choosing a^* may be approached by first considering Fig. 3.8. In this example, the value of $\eta(g_{2l})$ reaches a maximum at $a \approx 2.5$, with the performance being fairly insensitive to the exact value of a. In fact, a 50% change from $a=2.5$ yields less than a 9% change in efficacy. This suggests that a simple method may be used to find a nearly optimal estimate of a^*: First, arbitrarily choose three different breakpoints a_i for $i = 1,2,3$ and find the associated optimal tail slope b^*_i. Then evaluate $\eta_i = \eta(g_{2l})\Big|_{a_i;b^*_i}$ for $i = 1,2,3$ and fit a parabola through the three pairs of points (a_i,η_i). Finally, choose $\hat{a}^*$ as that point which maximizes the value of the fitted parabola.

Obviously, the initial choice of the three breakpoints cannot be completely arbitrary. The algorithm will perform best when the true value of a^* is bracketed by the values a_i, and $\eta(g_{2l})$ as a function of the breakpoint is approximately quadratic for $a_1 \le a \le a_3$. The use of this scaling procedure is demonstrated in Section 5.

4. Examples - Tail Matching Algorithm

We will now present examples of the use of g_{tm} in approximating some known optimal LO nonlinearities.

Generalized Gaussian Noise

The first comparison is between the approximate and exact versions of LO nonlinearities for the generalized Gaussian family. The exponent $\hat{c}$ is given by (3.7) after using the exact value c in (3.5) to obtain P_T. Since

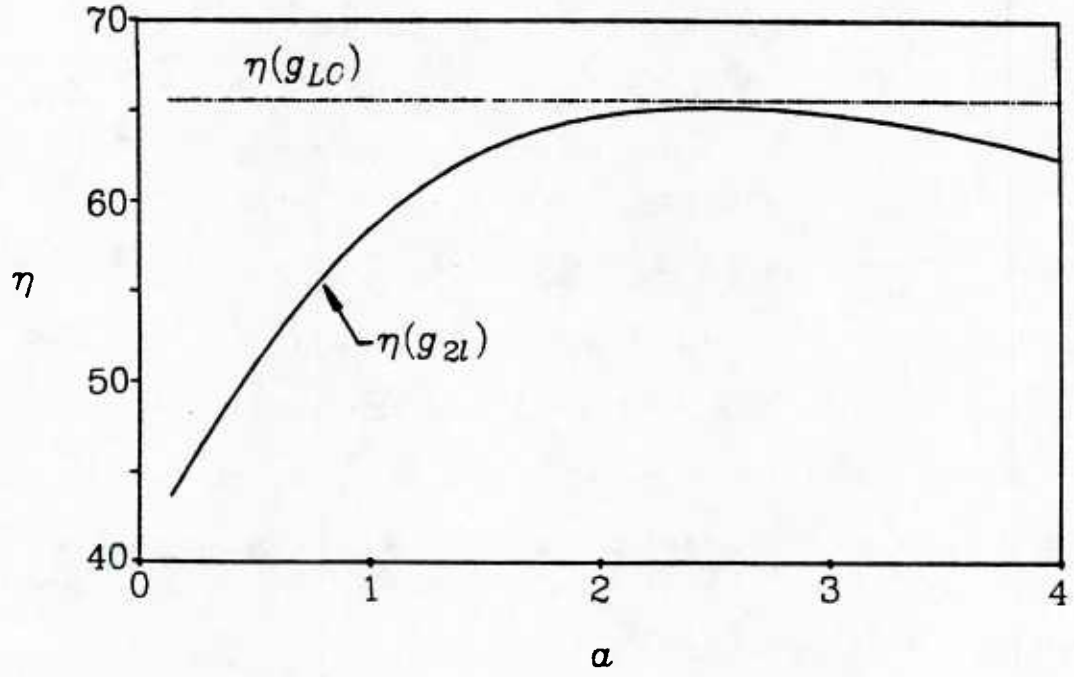

Fig. 3.8. Performance of g_{2l} for various breakpoints a with optimal tail slope b^*. A Gaussian-Gaussian ε-mixture density for the noise is assumed with $\varepsilon = .1$ and $\sigma_1^2/\sigma_0^2 = 750$.

this is an analytical example, and the true noise density is known, numerical methods can be used to obtain ν^*, the value of ν which maximizes $\eta(g_{tm})$. The performance of the suboptimal ZNL relative to the LO nonlinearity may be measured by asymptotic relative efficiency, given in terms of efficacy as

$$\mathrm{ARE}_{g_{tm},g_{LO}} = \frac{\eta(g_{tm})}{\eta(g_{LO})} \tag{3.16}$$

Figure 3.9 compares the performance of g_{tm}, the LO detector and a linear detector (ld), in terms of $\mathrm{ARE}_{g_{tm},g_{LO}}$ and $\mathrm{ARE}_{g_{LO},ld}$. The suboptimal nonlinearity performs quite well for the range $1 \le c \le 2$, but for $c < 1$, performance deteriorates. This is easily explained, since for small c, the LO nonlinearity output approaches $\pm\infty$ for inputs near zero, while the approximation method requires g_{tm} to pass through the origin

Johnson S_u Noise

Another family of heavy tailed densities introduced in the last chapter is the Johnson S_u family. The parameter δ controls tail heaviness, and the density has a Gaussian-like shape near the mode. For the purpose of example, it is a convenient density family to be used in studying the properties of the tail matching method. Since f_δ is given and known, $\hat{P}_T$ may be calculated from (3.5), and (3.7) gives $\hat{c}$. Again, numerical methods can be used to find the ν^* that maximizes efficacy. Some representative LO nonlinearities and suboptimal approximations are given for various values of δ in Fig. 3.10, and Figure 3.11 presents the performance comparison of g_{tm}, g_{LO}, and ld. For this family of densities, the

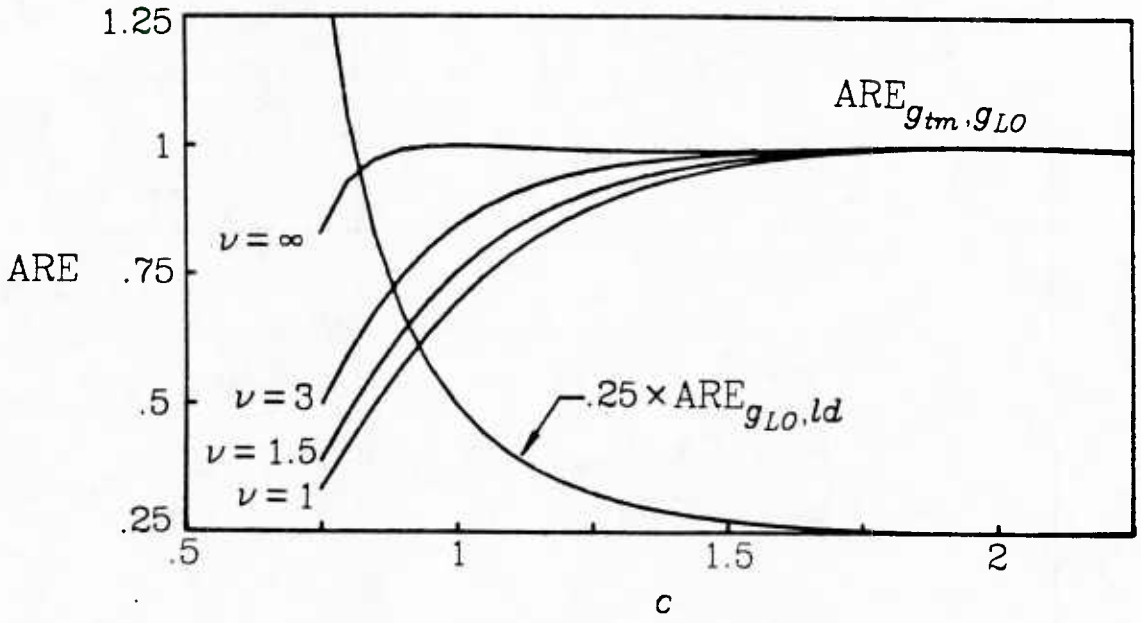

Fig. 3.9. Performance of g_{tm} and g_{LO} relative to the linear detector *ld* for various exponents c in the generalized Gaussian density.

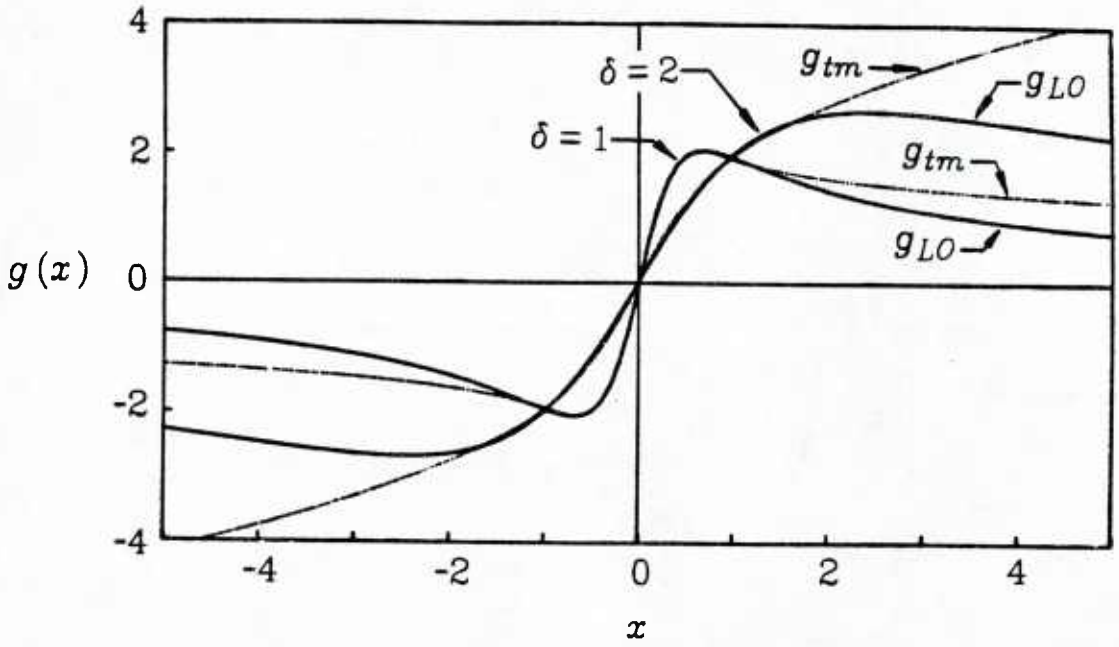

Fig. 3.10. The locally optimal nonlinearities g_{LO} and suboptimal nonlinearities for two members of the Johnson S_u family. The nonlinearity outputs are scaled for comparison purposes. For the case $\delta = 1$, the parameters of g_{tm} are $\hat{c} = .752$, $\nu^* = 3.26$, and $x_0 = 3$. For $\delta = 2$, $\hat{c} = 1.46$, $\nu^* = 1.88$, and $x_0 = 3$.

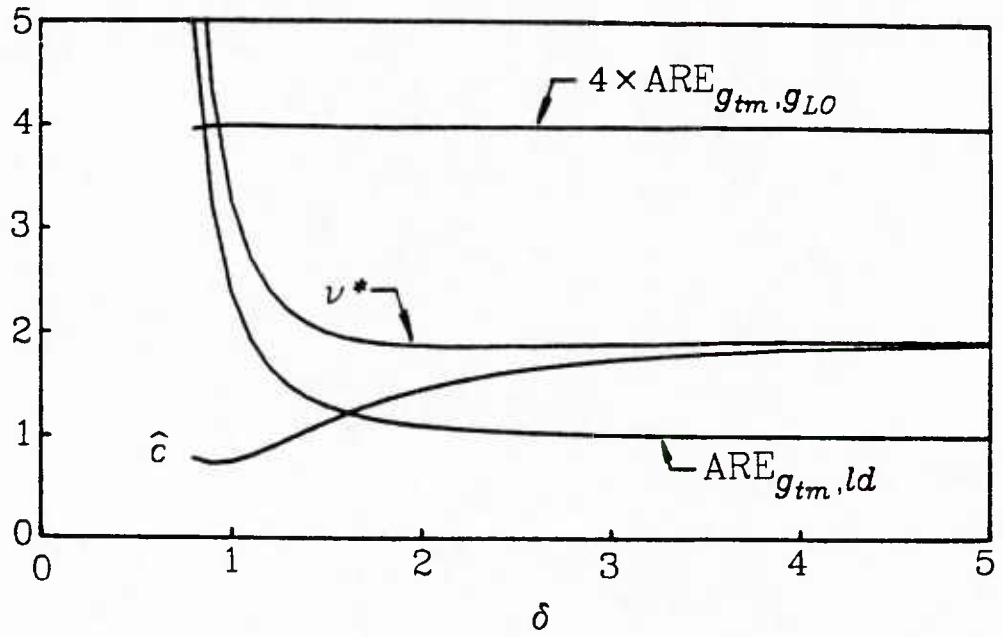

Fig. 3.11. Performance of nonlinearities g_{tm} and g_{LO} relative to the linear detector ld, for noise densities parameterized by δ of the Johnson S_u family. The optimal parameters $\hat{c}$ and ν^* are also given as a function of δ.

approximation method works quite well. Over the range $.8 \leq \delta \leq \infty$, the minimum of $ARE_{g_{tm}, g_{LO}}$ is .989, (occurring for $\delta = .8$). This means that only a small performance penalty would be incurred if g_{tm} were to replace the LO detector. As a final comment, it should be observed that, unlike the generalized Gaussian family, the Johnson S_u family fulfills the characteristics of a nearly Gaussian pdf given in Chapter 2, since f_c is sharply peaked, while f_δ has a Gaussian-like mode.

Gaussian-Gaussian ε-mixture Noise

The performance of g_{tm} in a third family of heavy tailed densities was also investigated. Here, the noise is assumed to be modeled by the ε-contaminated Gaussian-Gaussian mixture density, written as

$$f_\varepsilon(x) = (1-\varepsilon) f_0(x) + \varepsilon f_1(x) \tag{3.17}$$

where f_0 represents the pdf of a zero mean Gaussian random variable, and f_1 represents the pdf of another Gaussian random variable, with the variance ratio σ_1^2 / σ_0^2 large. The parameter ε controls the degree to which f_1 contaminates the nominal density f_0, and is typically taken to be small. Figure 3.12 shows a comparison between two LO nonlinearities and their corresponding approximations. The approximate nonlinearities g_{tm} do not appear as close to g_{LO} in this example as for the Johnson S_u family for two reasons: first, the tails of g_{LO} for the Gaussian-Gaussian ε-mixture increase almost linearly, while g_{tm} is constrained to have power law tails. Second, g_{LO} has a total of four local extrema, while g_{tm} is designed to have a maximum of two. On the other hand, g_{LO} for f_δ has two local extrema, and the tails asymptotically approach the x-axis.

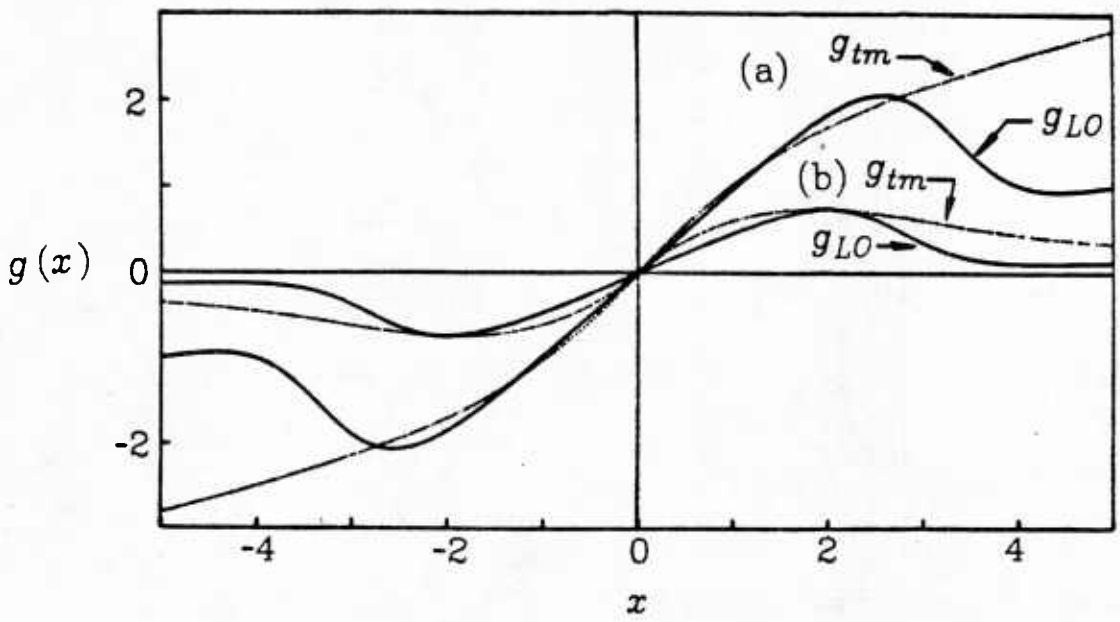

Fig. 3.12. The locally optimal nonlinearities g_{LO} and suboptimal nonlinearities g_{tm} for two members of the Gaussian-Gaussian ε-mixture family. (a) $\varepsilon = .05$, $\sigma_1^2/\sigma_0^2 = 5$, $\hat{c} = 1.54$, $\nu^* = .957$, $x_0 = 3$. (b) $\varepsilon = .20$, $\sigma_1^2/\sigma_0^2 = 20$, $\hat{c} = -.196$, $\nu^* = .821$, $x_0 = 3$.

The performance of g_{tm} was computed for a range of values of ε and σ_1^2/σ_0^2 suggested by Vastola [18,19] as being representative of physical noise situations. Figures 3.13 and 3.14 present the performance of the tail matching method. The sets of curves indicate that the tail matching algorithm generates nonlinearities which work quite well relative to the optimal detector in Gaussian-Gaussian ε-mixture noise.

The results show that it is often possible to achieve nearly optimal performance using this simple approximation method. The salient feature of the noise tail matching method is its ability to adjust tail behavior in accordance with simple observations of the noise tail heaviness.

Simulation

To see how well this system might work in practice, some actual physical noise was used to drive the system. The noise was collected underneath the Arctic ice pack, and details may be found in [22]. A summary of the data selected for simulation purposes is given in Appendix 2.1 of this thesis. The noise data is highly nonstationary; a background Gaussian noise is abruptly interrupted with segments of a high variance noise generated during cracking of the ice pack.

To get a more nearly stationary noise for driving the system, the data in each block was adjusted to zero mean and randomly permuted, thereby simulating the output of a stationary noise source. This adjustment was necessary solely to improve the rate of convergence of the stochastic approximation algorithm for obtaining ν^*. Figures 3.15 and 3.16

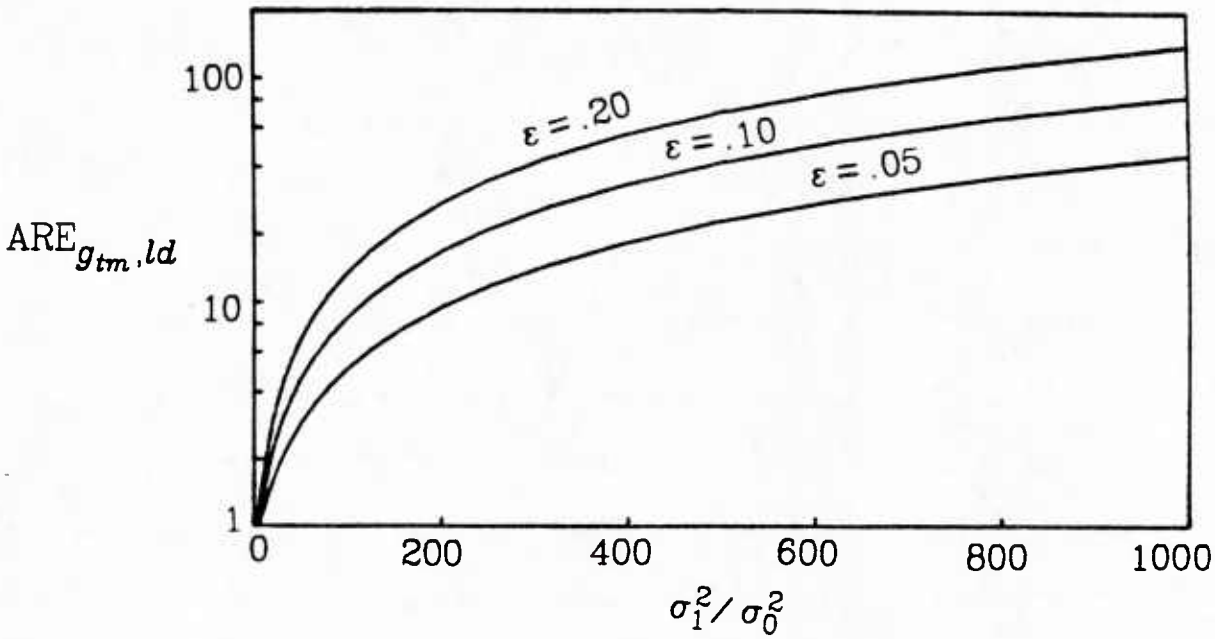

Fig. 3.13. Performance of g_{tm} relative to the linear detector *ld* in Gaussian-Gaussian ε-mixture noise for various values of ε and a range of variance ratios σ_1^2/σ_0^2.

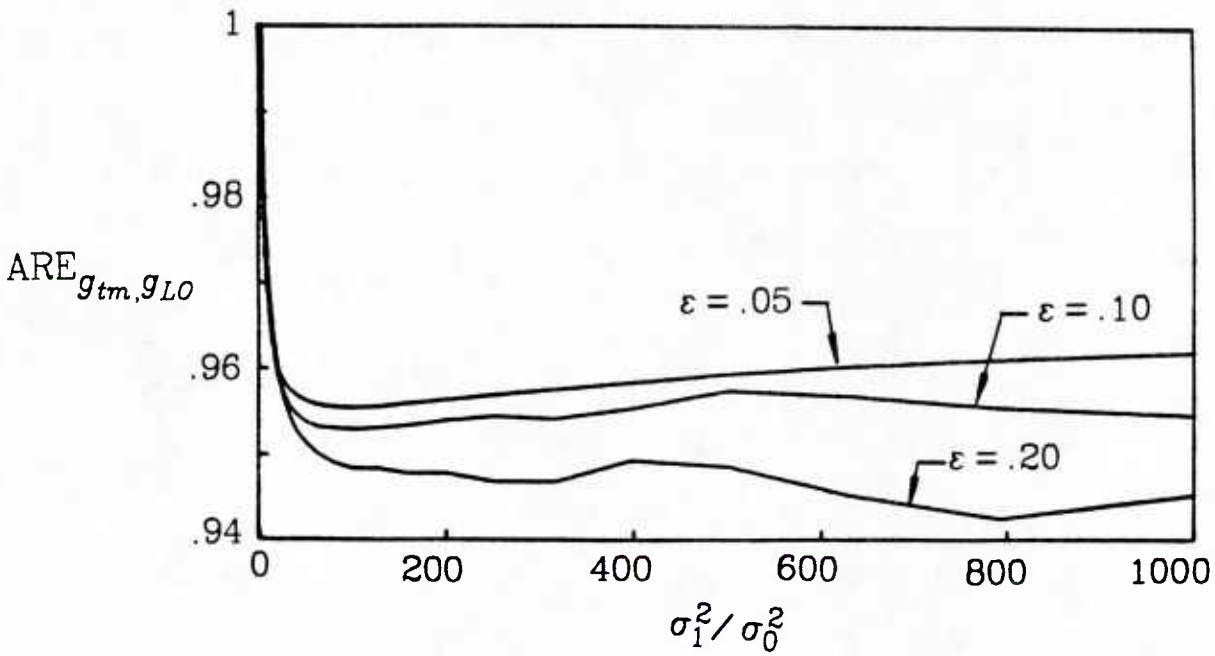

Fig. 3.14. Performance of g_{tm} relative to g_{LO} in Gaussian-Gaussian ε-mixture noise for various values of ε and range of variance ratios σ_1^2/σ_0^2. Curves are approximate due to numerical roundoff errors.

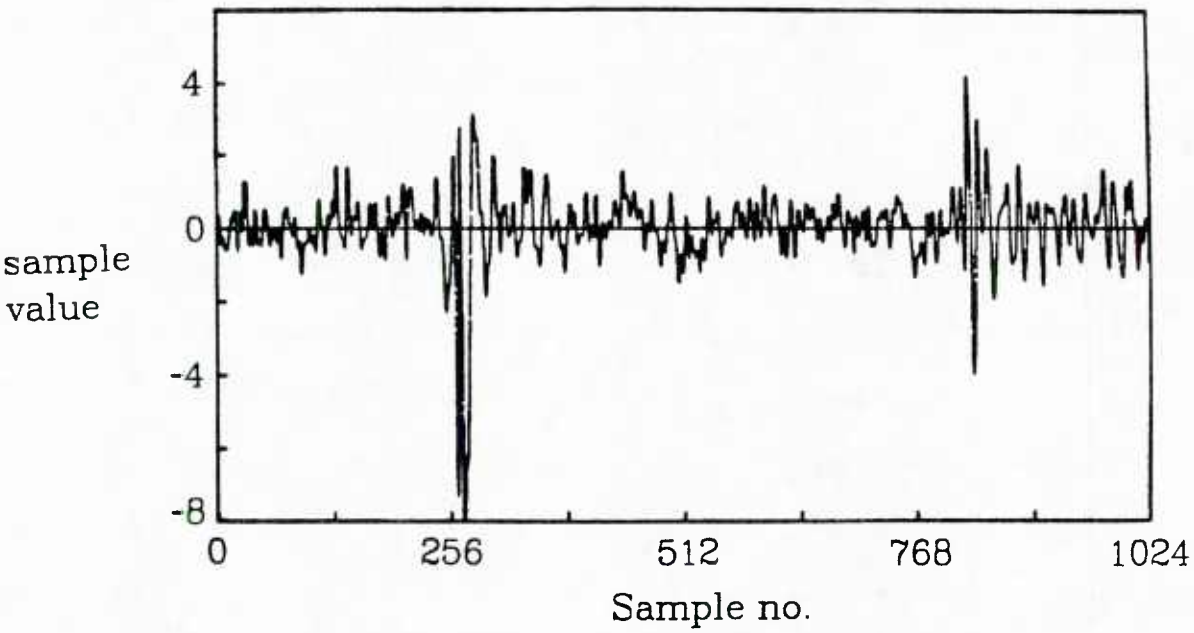

Fig. 3.15. Sample Arctic under-ice noise data, record 2220, adjusted to zero mean. Vertical scale is in standard deviations from the mean.

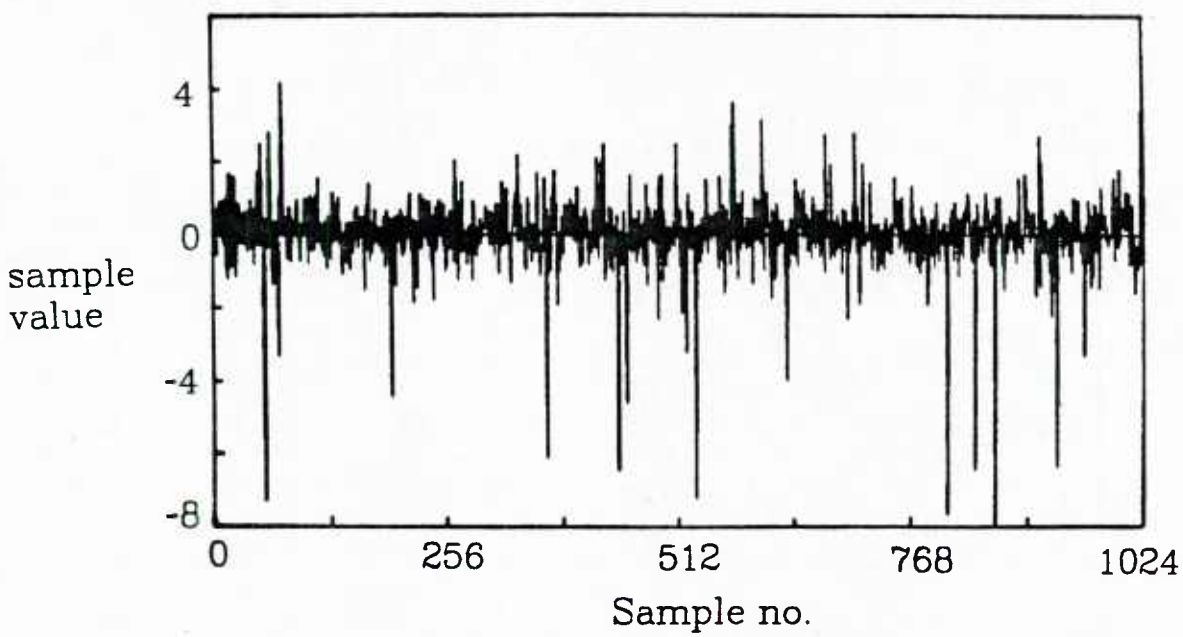

Fig. 3.16. Sample Arctic under-ice noise data, record 2220, adjusted to zero mean, and randomly permuted. Vertical scale is in standard deviations from the mean.

present a sample block of data, before and after random permutation, respectively.

This noise was used as the input for the tail matching algorithm. A threshold $T = 3\sigma$ was chosen for estimating the tail probability of the noise, where σ is the standard deviation of noise data block. As more noise data is observed, the cumulative tail probability estimate converges to the true tail probability. The exponent $\hat{c}$ was estimated from this cumulative estimate of P_T. The simulated system had no knowledge of the true density generating the noise observations; therefore, the Kiefer-Wolfowitz stochastic approximation method was used to find the value of ν^* which maximized $\eta(g_{tm})$ for $x_0 = 3$. The convergence rate towards ν^* is fixed by the particulars of the stochastic approximation algorithm, and no formal attempts were made to optimize its performance.

Figure 3.17 shows the running estimate of $\hat{c}$ and ν as a function of sample number, and Fig. 3.18 shows the estimated value of $\mathrm{ARE}_{g_{tm},ld}$ for each block of 1024 samples. Since the true distribution of the noise is unknown, $\eta(g_{tm})$ was calculated by evaluating (3.9) using the empirical distribution of the data block under consideration and the current estimate of g_{tm}. The estimate of ARE results when $\eta(g_{tm})$ is multiplied by the variance of the noise data block.

At the end of the simulation, it was assumed that the parameters of g_{tm} were as near optimal as possible. These final values are given in Fig. 3.19. It was desired to compare the performance of g_{tm} to the performance of the linear detector. To do this, (3.9) was evaluated using the final estimate of g_{tm} and the empirical distribution of the 58 blocks of

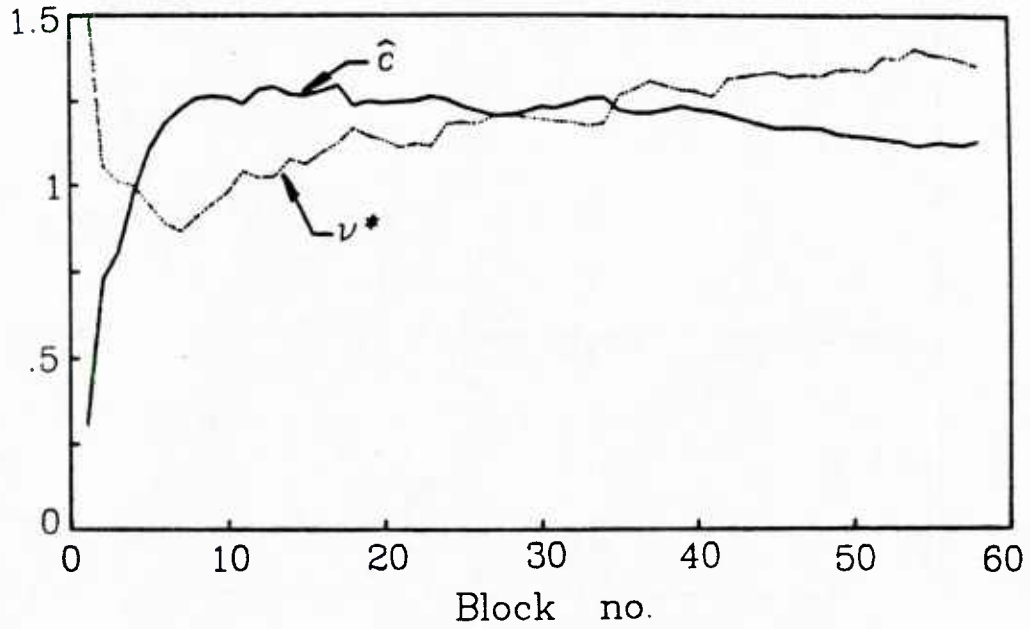

Fig. 3.17. Parameters $\hat{c}$ and ν^* for each selected Arctic under-ice noise block of 1024 samples.

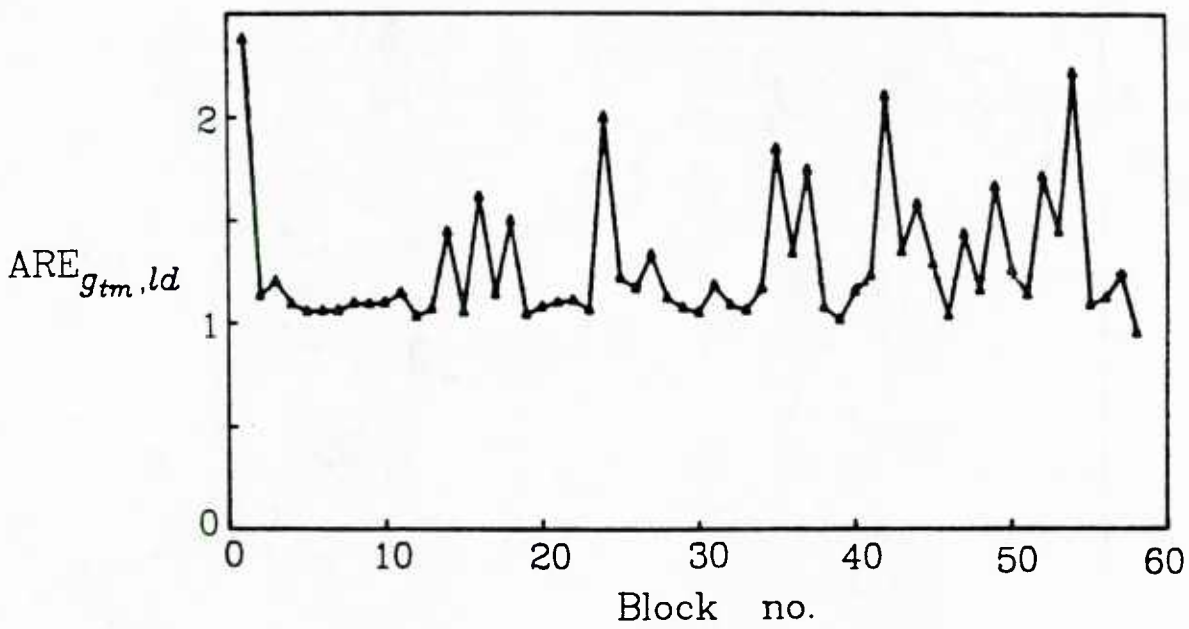

Fig. 3.18. The estimated performance of g_{tm} relative to the linear detector for each selected Arctic under-ice noise data block. The parameters of g_{tm} are those given in Fig. 3.17.

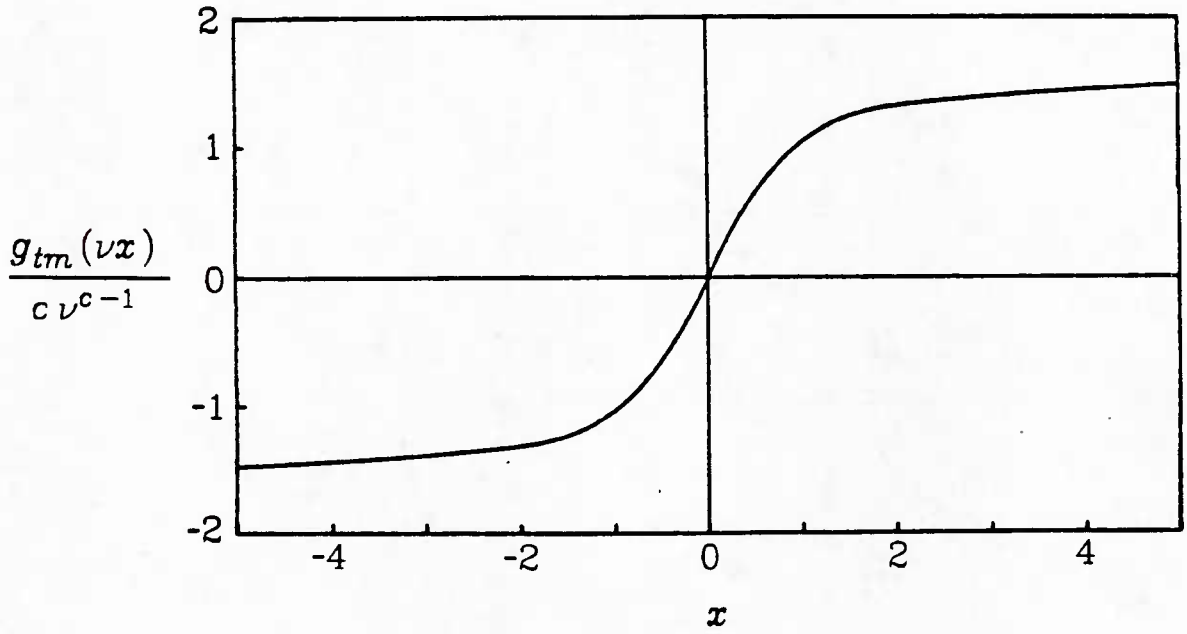

Fig. 3.19. The final estimated nonlinearity g_{tm} at end of Arctic under-ice noise data simulation. The final parameters are $\hat{c} = 1.13$, $\nu^* = 1.35$, $x_0 = 3$.

1024 noise samples, and the result was multiplied by the variance of the entire data set, yielding $ARE_{g_{tm},ld} = 1.39$ as a performance estimate.

Because the true distribution is unknown, g_{LO} cannot be found, and it is not possible to calculate the performance of the LO detector. However, it is possible to conclude that the tail matching procedure was able to adapt the suboptimal ZNL in a constructive way, for g_{tm} shows improved performance over the linear detector.

5. Examples - Efficacy Maximization Algorithm

Laplace Noise

Figure 3.20 provides a representative example of g_{2l} when its parameters are estimated assuming a Laplace density for the noise. The LO detector in this case is a sign detector, sd. Intuition might suggest that the best approximation employing two linear regions and fixed nonzero breakpoint is the amplifier-limiter $al(x;a>0) = g_{2l}(x;a>0,b=0)/a$, but this turns out not to be the case. The best performance is obtained when g_{2l} has tails that return to meet the x-axis. Fig. 3.21 compares the performance of g_{2l}, sd, and al in terms of their ARE relative to the linear detector ld. Note that g_{2l} has improved performance over both the linear detector and the amplifier-limiter detector for any choice of $a \neq 0$. Also, when $a \to 0$, both al and g_{2l}/a approach the form of sd, and their performances converge towards the optimal performance of sd.

For each particular value of a, the optimal tail slope b^* was found by iterative application of (A3.6) and (3.2). Convergence was typically rapid, often requiring 3 iterations for a change of less than .001 in b^*.

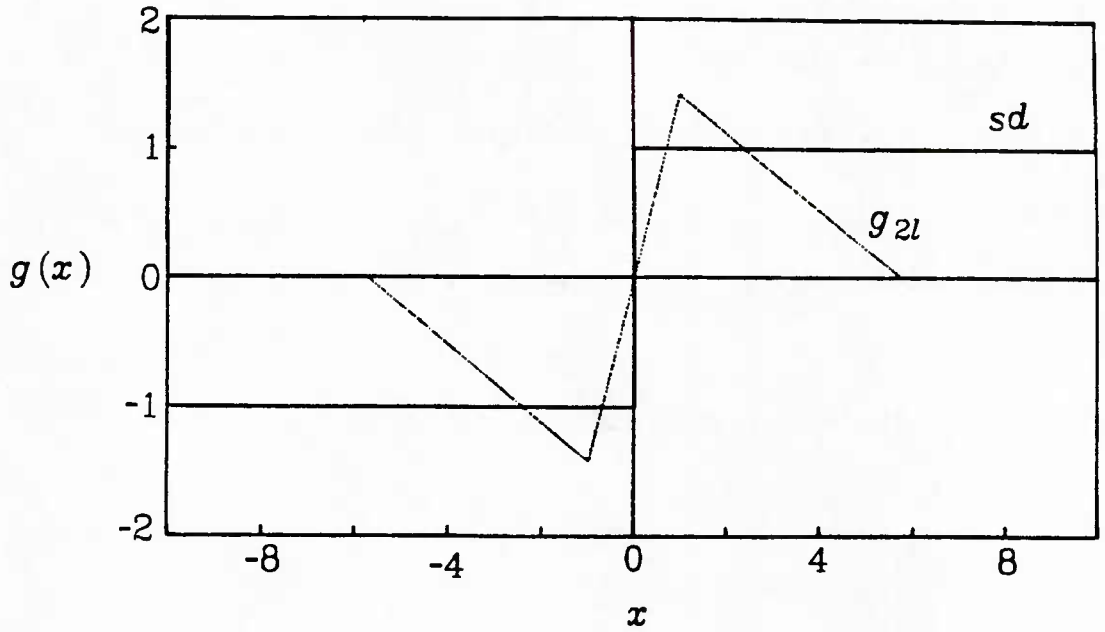

Fig. 3.20. Comparison of the nonlinearity g_{2l} and the sign detector sd for $a = 1$ and $b^* = -.211$. The output of g_{2l} is scale for comparison with sd.

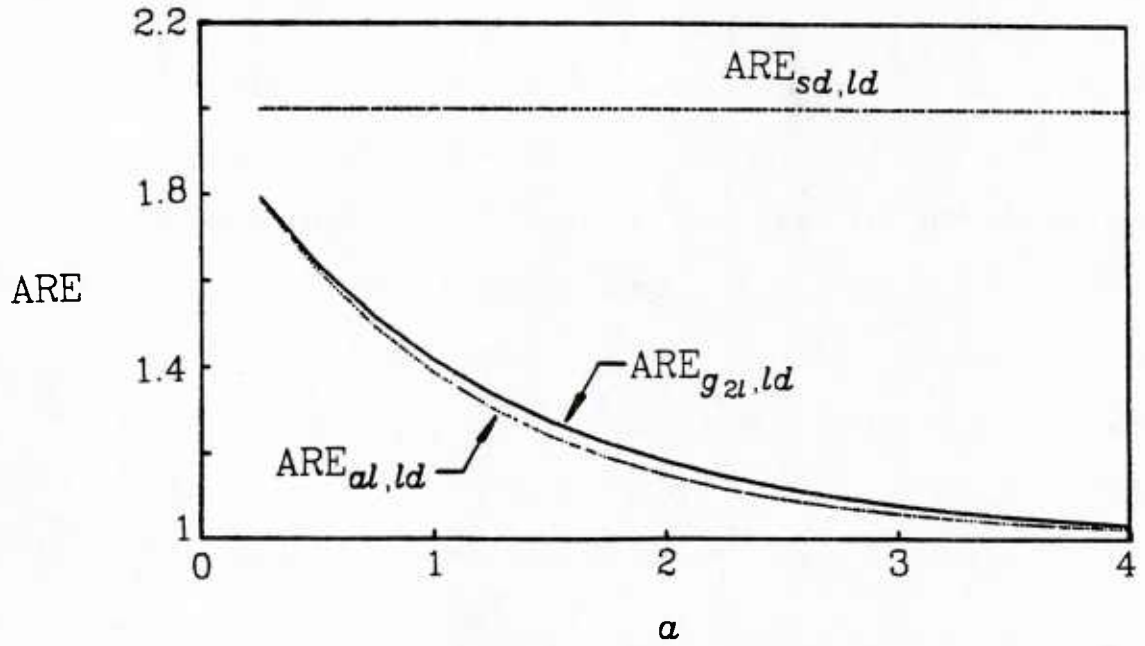

Fig. 3.21. Performance of the nonlinearity g_{2l}, the amplifier limitier al, and the sign detector sd relative to the linear detector ld for Laplace noise and various breakpoints a. Tail slope b^* is optimal for each choice of a.

Gaussian-Gaussian ε-mixture Noise

Here we consider the performance of g_{2l} in the presence of Gaussian-Gaussian ε-mixture noise. A representative example of g_{LO} and g_{2l} is given in Fig. 3.22 for f_ε with parameters chosen in the middle of the range suggested by Vastola [18,19] as being reasonable for observed cases of Middleton's Class A noise density.

At the end of Section 3 a technique was described for finding the estimated optimal breakpoint $\hat{a}^*$. It was employed for this example by choosing the 3 points $a_1 = .5q_{.9}$; $a_2 = q_{.9}$; and $a_3 = 2q_{.9}$; with $q_{.9}$ meaning the .9 quantile of the distribution. In practice, these quantities are easily measured characteristics of a noise distribution. For the particular example of Gaussian-Gaussian ε-mixture noise, it was found that $\hat{a}^*$ was typically within 5% of the true value of a^*, and the efficacy of g_{2l} using $\hat{a}^*$ was within 1% of the maximum possible efficacy of g_{2l}. Further, the estimate $\hat{a}^*$ was stable for different choices of $\{a_1,a_2,a_3\}$. Note that for this example the true values of a^* were available only through computationally burdensome numerical methods.

Given the estimated optimal breakpoint $\hat{a}^*$, the slope b^* was found iteratively, as before. Convergence of b^* to within a .001 change occurred typically within 6 iterations. Figure 3.23 shows $\mathrm{ARE}_{g_{2l},ld}$ for various combinations of ε and σ_1^2. Figure 3.24 compares the performance of g_{2l} and g_{LO}, where it may be seen that the performance of g_{2l} is within a few percent of the optimal, and at worst, within 4%. The relatively

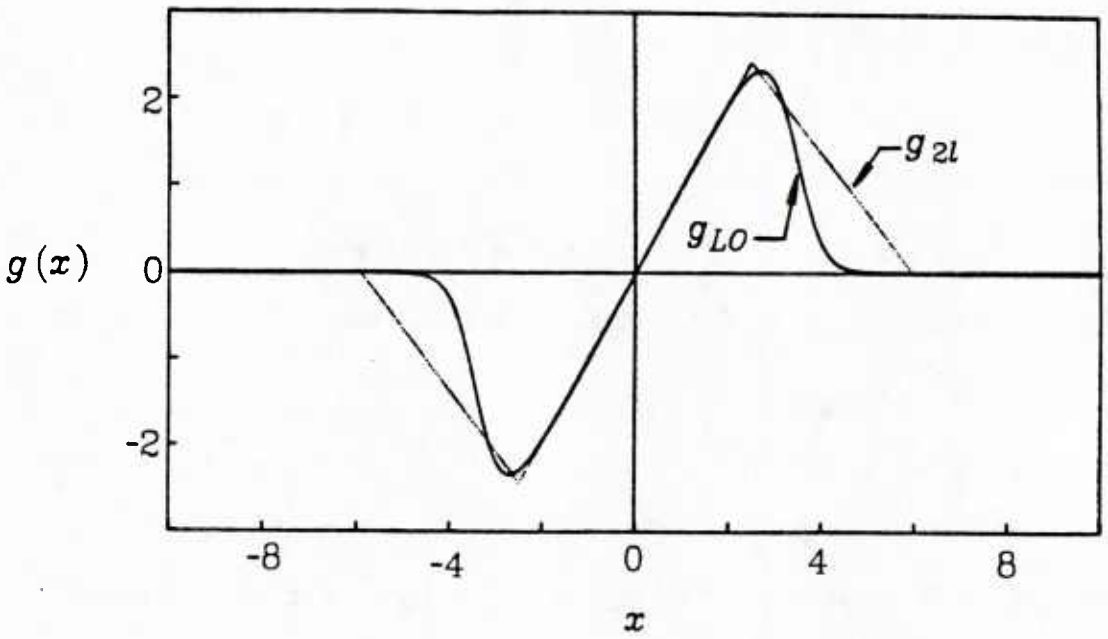

Fig. 3.22. Representative nonlinearities g_{2l} and g_{LO} for Gaussian-Gaussian ε-mixture noise with $\varepsilon = .1$, $\sigma_1^2/\sigma_0^2 = 750$, $a = 2.5$ and $b^* = -.524$. The output of g_{2l} is scaled for comparison.

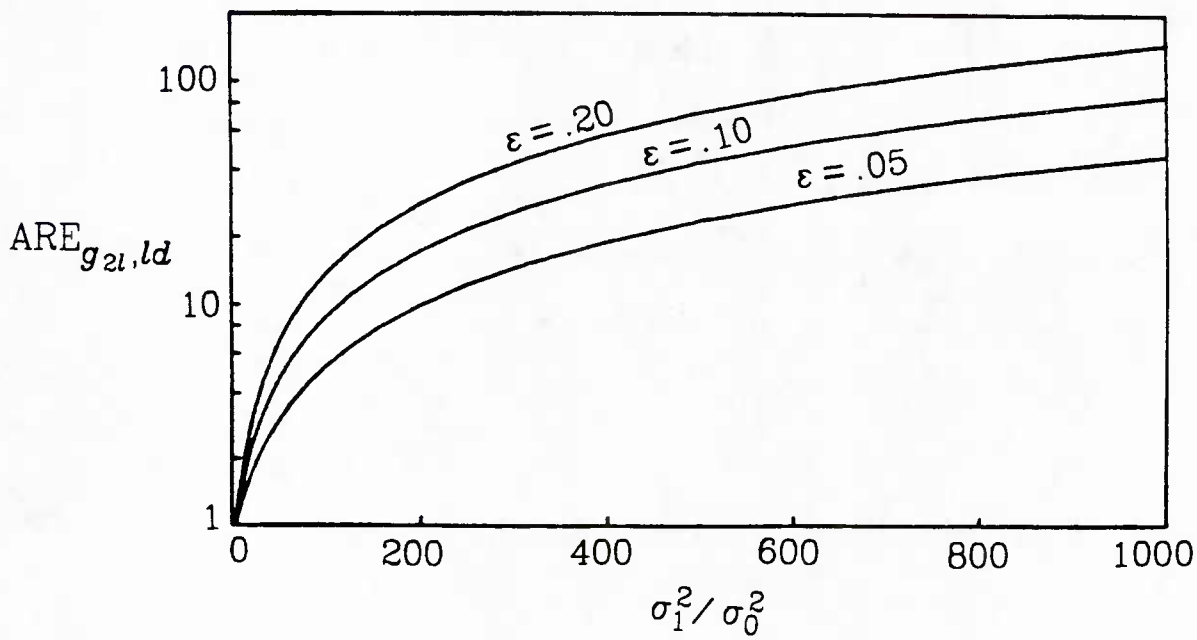

Fig. 3.23. Performance of the nonlinearity g_{2l} relative to the linear detector *ld* in Gaussian-Gaussian ε-mixture noise for various values of ε and range of variance ratios σ_1^2/σ_0^2.

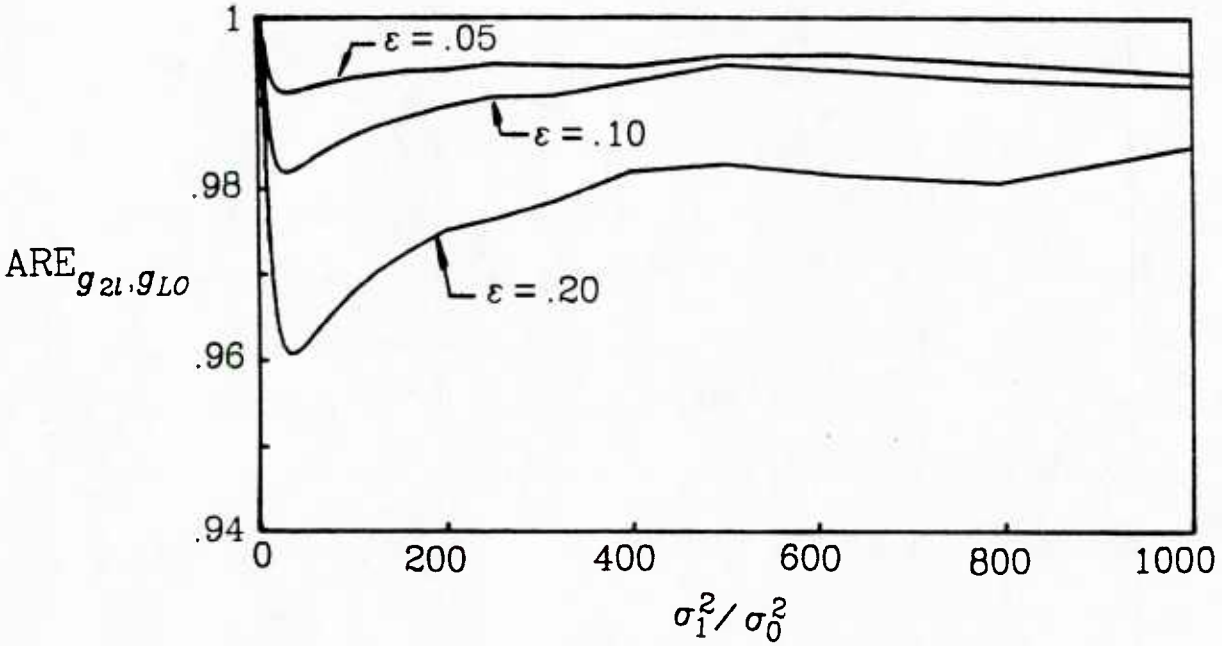

Fig. 3.24. Performance on the nonlinearity g_{2l} relative to the locally optimal nonlinearity g_{LO} for Gaussian-Gaussian ε-mixture noise, for various values of ε and range of variance ratios σ_1^2 / σ_0^2.

poorer performance occurs for the larger values of ε which assign more probability mass to a region away from the origin. There the tails of the LO nonlinearity diverge from the x-axis, while the approximate nonlinearity g_{2l} has truncated tails in this region. As a result, its performance suffers slightly.

Simulation

As was done for the tail matching algorithm, the Arctic under-ice noise data was used to examine the performance of the efficacy maximizing procedure. The same 58 high-kurtosis data blocks used previously and described in Appendix 2.1 were used to drive the algorithm, after each data block was adjusted to zero mean. Unlike the simulation in Section 4, no further manipulation was necessary to prepare the data.

Here the .9 quantile of the noise distribution was estimated for each data block as

$$\hat{q}_{.9} = \frac{q_{.9} - q_{.1}}{2} \tag{3.18}$$

to minimize the effects of the high skew occasionally observed. The 3-point parabolic fitting method was used to estimate a^*, with $a_1 = .5\hat{q}_{.9}$; $\hat{q}_{.9} = a_2$; and $2\hat{q}_{.9} = a_3$ serving as the three arbitrary breakpoints. A minor modification to the algorithm was made, requiring that $a_1 \le \hat{a}^* \le a_3$. Any $\hat{a}^*$ outside this range was replaced by a_1 or a_3, as appropriate. The modification ensures that the algorithm does not produce highly inaccurate values of $\hat{a}^*$ when the interval $[a_1, a_3]$ does not bracket the true value a^*. The estimated values $\hat{a}^*$, and ultimately, the performance of g_{2l} were insensitive to using $\hat{q}_{.85}$ or $\hat{q}_{.95}$ instead of $\hat{q}_{.9}$.

Once $\hat{a}^*$ was found for each block, b^* was found using the previously discussed iterative procedure. Figure 3.25 shows the estimated values $\hat{a}^*$ and b^* for each of the data blocks in the simulation. Note that both of the parameters appear to have fairly steady nominal values. For each data block, $\eta(g_{2l})$ was estimated by taking the current value of $\hat{a}^*$ and b^* and evaluating (3.14) or (3.15), as appropriate, with respect to the noise data block empirical distribution. Multiplying this result by the estimated variance of the data block yields an estimate of $ARE_{g_{2l},ld}$, shown in Fig. 3.26.

At the end of the simulation, the average values taken by a^* and b^* were computed, and are given in Fig. 3.27 along with a depiction of g_{2l} using the average values. Again, if these "final" parameter values are used in (3.15) for the entire 58 blocks of 1024 noise samples, and the result is multiplied by the overall noise variance, $ARE_{g_{2l},ld} = 1.39$ is obtained as an estimate of performance. Surprisingly, this is exactly the same result as the tail matching algorithm overall performance.

6. Conclusion

The conclusion to be drawn from this study is that it is possible to implement adaptive detector nonlinearities using fairly simple techniques.

Tail Matching

Of the two methods suggested, the first, utilizing an estimate of tail behavior, is quite simple: let the tails of the suboptimal ZNL be the tails of the locally optimal nonlinearity for a density with the same tail

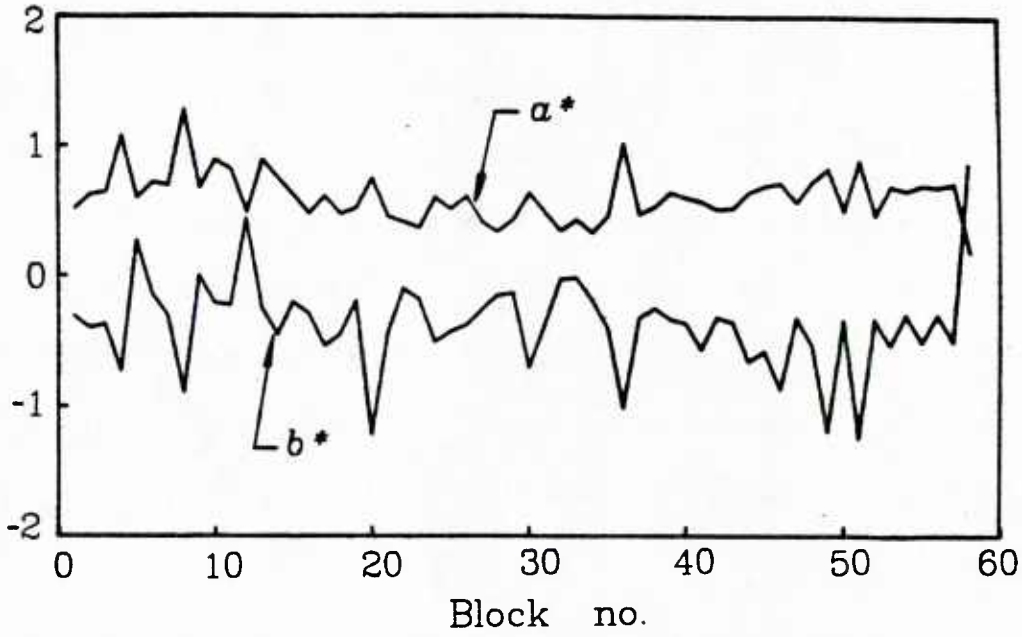

Fig. 3.25. Estimated parameters a^* and b^* of the nonlinearity g_{2l} for each of the selected Arctic under-ice noise data blocks.

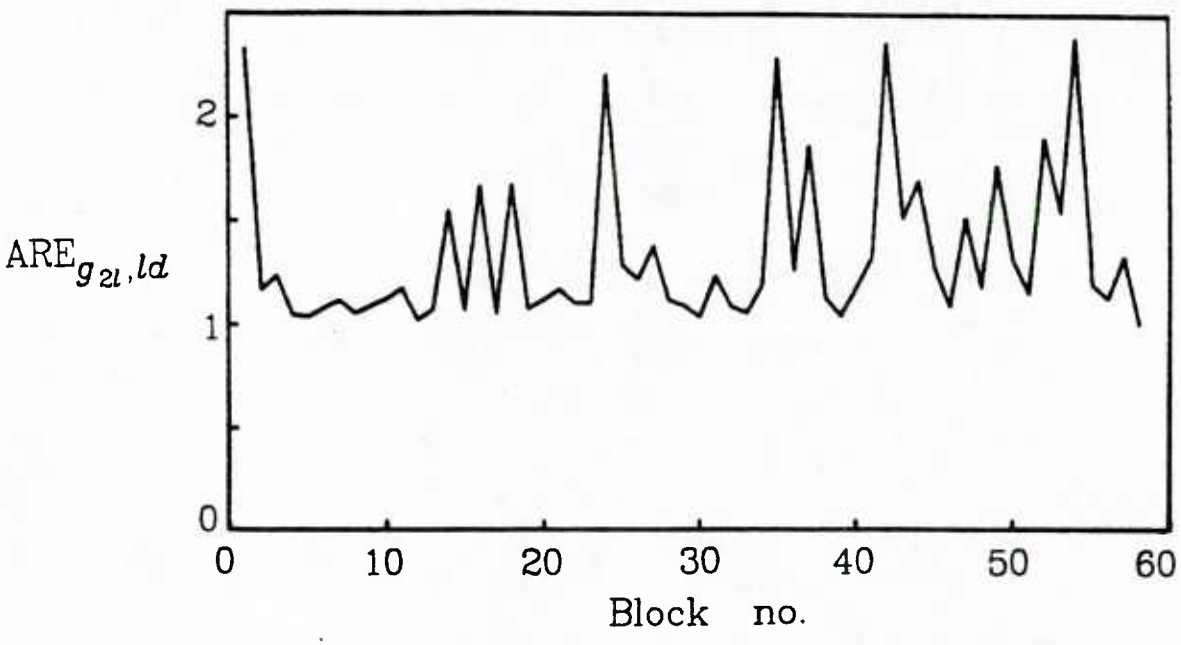

Fig. 3.26. The estimated performance of g_{2l} relative to the linear detector ld for each of the selected Arctic under-ice noise data blocks. The parameters of g_{2l} are those given in Fig. 3.25.

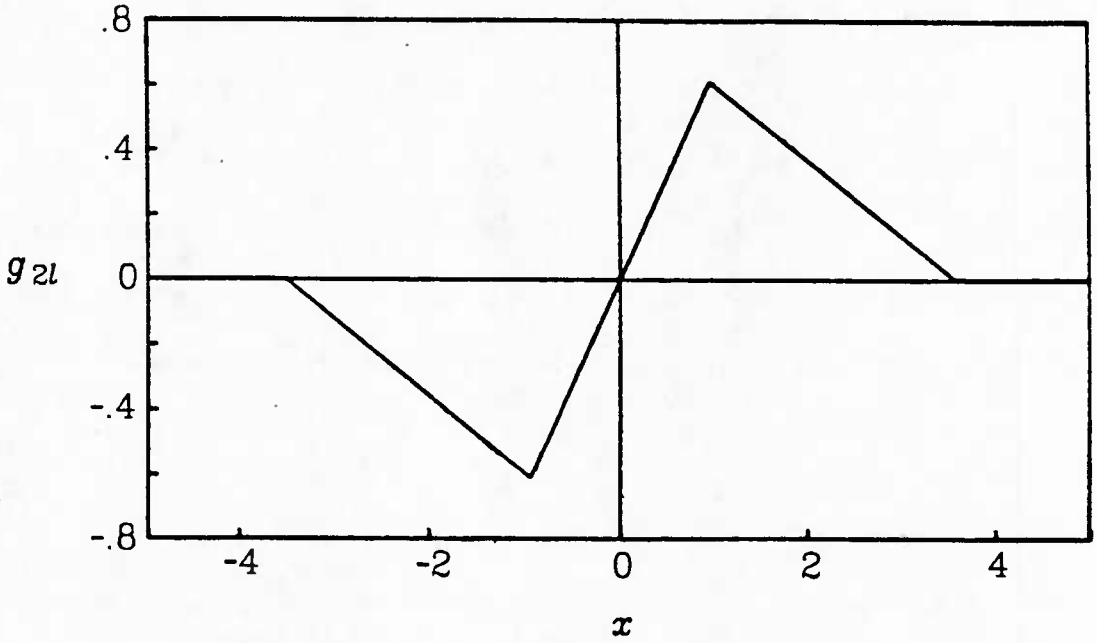

Fig. 3.27. The nonlinearity g_{2l} for the average parameters $a^* = .610$ and $b^* = -.372$. These values are the average of the values given in Fig. 3.25.

probability mass as the observed noise distribution. Apparently P_T conveys enough information about tail behavior of the noise density and fairly good performance results using even a crude approximation to the true ZNL tails. A more sophisticated estimate of $\hat{c}$ might improve the performance g_{tm}, as might a different choice altogether for the class of density tails used for g_{tm}. It would be interesting to discover how much additional complexity any resulting performance gain could justify.

Other related approaches were recently explored by Modugno [17]. Some work done by Miller and Thomas [9] in approximation of LO nonlinearities suggests that even very simple approximants of the optimal nonlinearity have the potential to achieve performance which is acceptably near the optimal.

Efficacy Maximization

The second method suggests maximizing the performance of a simple generic nonlinearity. The approximate detector ZNL consists of a central linear region of unity slope surrounded by linear tail regions, generally of different slope. A closed form expression (A3.6) for the tail slope is given, and a method is suggested for determining approximately the appropriate breakpoint between central and tail regions in the ZNL. The tail slope is either known exactly after a single application of (A3.6), or after a few iterations using (A3.6).

Some examples show that the performance of the suboptimal ZNL compares well with the LO detector performance, at least when the partial moments of the noise density are known exactly. In the examples,

the performance is very good, usually within a few percent of the optimal performance. In practice, the performance of g_{2l} may not be quite so good, since at best only estimates of the partial moments would be available. However, these partial moments are easily estimated, since the highest order is second degree. Also, each integration typically spans a region containing a nontrivial amount of probability mass; therefore it should be fairly easy to converge quickly to low variance estimates of the partial moments. The issue of sensitivity of $\eta(g_{2l})$ as a function of errors in the partial moments has not been examined at this time.

The advantage of this method is that implementation of the proposed nonlinearity is quite simple: all that is required is the ability to apply different linear gains (plus constant offsets) to inputs occurring along different regions of the x-axis. As a result, adaptation of g_{2l} can be accomplished with little overhead, once a^* and b^* are known.

The chief disadvantage of this method is the fact that negative values of b^* must be found iteratively. However, intermediate values of b^* are useful; the performance of g_{2l} is not maximal, but it is nearly so, and the performance improves monotonically with each iteration.

Another complication is the fact that an explicit solution for a^* is not available. The parabolic fitting method mentioned may be used to estimate a^*, or other methods may be used to converge to the best value. On the other hand, it appears that precise placing of the breakpoint is not a critical matter.

The parabolic fitting procedure for finding the breakpoint is also applicable to finding the appropriate scale factor to the input of a ZNL.

Instead of performing a Kiefer-Wolfowitz stochastic approximation to find the optimal scale factor, the parabolic fitting method could be adapted for solving the scaling problem. The quantities involved in the parabolic fitting method are expectations of the noise observations transformed by the square or first derivative of the ZNL. For heavy-tailed noises, the nonlinearity tails allow the very large noise observations to have much less influence than if they were untransformed or linearly processed; therefore, the large observations contribute very little to the computed expectation of the square and first derivative of the nonlinearity. Intuition suggests that this inherently might be a more robust procedure than computing variance through expectation of the squared noise observations.

Comparison of Algorithms

The performance of g_{tm} and g_{2l} may be compared by computing $\text{ARE}_{g_{2l},g_{tm}}$ under identical noise situations. Because of the appeal of f_ε as a reasonable model for certain observed noise densities, it will be used as the standard for comparison. Figure 3.28 presents $\text{ARE}_{g_{2l},g_{tm}}$, where it may be seen that there is some advantage to the efficacy maximizing algorithm giving g_{2l}. This should not be surprising, since the algorithm for g_{tm} is an "open loop" procedure which does not optimize the ZNL shape. (Both algorithms optimize scale with respect to efficacy.) Further, g_{2l} may be regarded as having simpler shape than g_{tm}, for it is piecewise linear, while g_{tm} has power law tails and a polynomial central region.

As was noted at the end of the simulations, the estimated perfor-

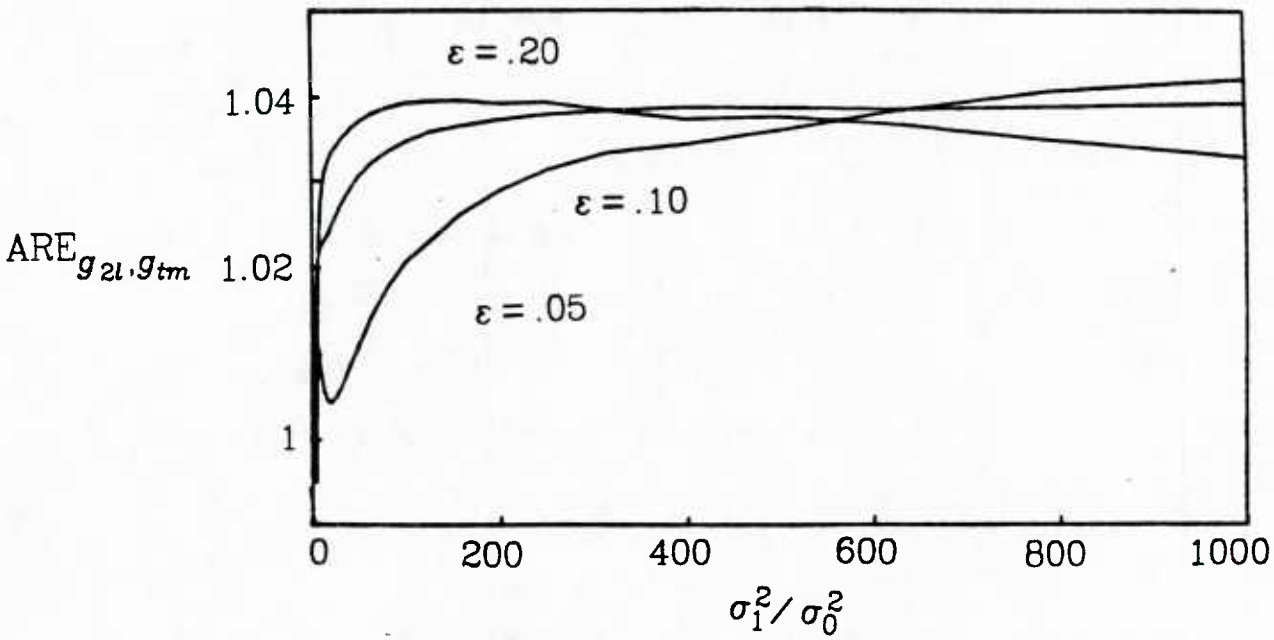

Fig. 3.28. Performance of the nonlinearity g_{2l} relative to the nonlinearity g_{tm} in Gaussian-Gaussian ε-mixture noise for various ε and range of variance ratios σ_1^2/σ_0^2.

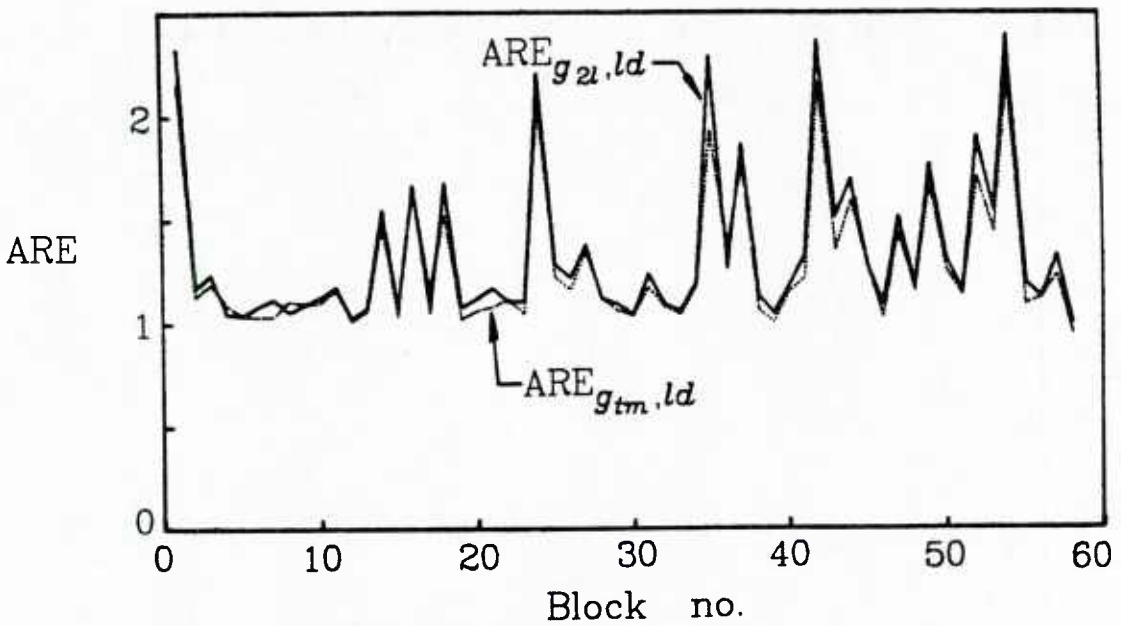

Fig. 3.29. The estimated performance of the nonlinearities g_{2l} (solid line) and g_{tm} (broken line) relative to the linear detector *ld* for each of the selected Arctic under-ice noise data blocks. The parameters of the nonlinearities are the values current for each noise block.

mance improvement was 1.39 for either detector relative to a linear detector. However, the value of $ARE_{g_{2l},ld}$ was obtained for a single fixed pair of parameters $(\hat{a}^*,b^*)$. If $\hat{a}^*$ and b^* are allowed to vary as the noise statistics change from block to block, Fig. 3.29 illustrates that g_{2l} may have a slight advantage over g_{tm}. The reason for this is that, while the adaptation algorithm for g_{tm} forces the parameters to converge, the algorithm for g_{2l} does not include any memory of the parameters from the previous noise data block. Thus, the adaptation of g_{2l} may be considered as more agile.

Both algorithms, to some extent, are *ad hoc*. The purpose in exploring these methods was not to find a definitive algorithm for designing nearly optimal, but simple, detector nonlinearities. Instead, the objective was to gain insight into this problem. The conclusion is that even relatively unsophisticated ZNL design techniques have potential for highly successful application, provided the design algorithm has available some knowledge of the noise density tail shape.

Appendix 3.1

We seek to maximize (3.14) with respect to b, and begin with some notational preliminaries. The first step is to expand the denominator to obtain

$$\eta = \frac{2\left[\int_0^a f + b\int_a^\infty f\right]^2}{\int_0^a x^2 f + b^2\int_a^\infty x^2 f + 2ab(1-b)\int_a^\infty xf + a^2(1-b)^2\int_a^\infty f} \qquad \text{(A3.1)}$$

For convenience, we rewrite (A3.1) as

$$\eta = \frac{2[I_1+bI_2]^2}{I_3+b^2I_4+2ab(1-b)I_5+a^2(1-b)^2I_2} \qquad \text{(A3.2)}$$

Taking the partial derivative of (A3.2) with respect to b, we obtain

$$\frac{\partial\eta}{\partial b} = \left[\frac{4(I_1+bI_2)}{d^2}\right] \times \left[\begin{matrix} b(aI_2I_5-a^2I_2^2-I_1I_4+2aI_1I_5-a^2I_1I_2) \\ +(I_2I_3+a^2I_2^2-aI_1I_5+a^2I_1I_2)\end{matrix}\right] \qquad \text{(A3.3)}$$

with d representing the denominator of (A3.2). The notation is further simplified by rewriting (A3.3) as

$$\frac{\partial\eta}{\partial b} = \left[\frac{4(I_1+bI_2)}{d^2}\right]\left[bC+D\right] \qquad \text{(A3.4)}$$

A necessary condition for (A3.1) to have a maximum is that $\frac{\partial\eta}{\partial b} = 0$ has a solution for some b. A solution always exists, since the roots of (A3.4) are the pair $\{(-I_1/I_2),(-D/C)\} = \{b_0,b_1\}$.

A sufficient condition for a maximum to occur at $b^*\in\{b_0,b_1\}$ is $\left.\frac{\partial^2\eta}{\partial b^2}\right|_{b^*} < 0$. It is not necessary to evaluate the second derivative; all

that is required is to note that the numerator of the first derivative is quadratic in b, and the denominator is positive for all b. Although $\frac{\partial \eta}{\partial b}$ is not parabolic, $\frac{\partial^2 \eta}{\partial b^2}$ has the same sign at b^* as the slope of the quadratic numerator evaluated at b^*, one of its roots. Therefore, the root of (A3.3) giving maximum efficacy will be located on the negative sloping branch of the numerator parabola. Taking this into account, and the fact that, if $C > 0$, the smaller root is on the negative sloping branch, and if $C < 0$, the larger root is on the negative sloping branch, the solution of (A3.4) which maximizes (A3.1) is

$$b^* = \begin{cases} \max\{b_0, b_1\} & \text{for } C < 0 \\ \min\{b_0, b_1\} & \text{for } C > 0 \end{cases} \tag{A3.6}$$

where

$$b_0 = -I_1 / I_2$$

$$b_1 = -\frac{I_2 I_3 + a^2 I_2^2 - a I_1 I_5 + a^2 I_1 I_2}{a I_2 I_5 - a^2 I_2^2 - I_1 I_4 + 2a I_1 I_5 - a^2 I_1 I_2}$$

and

$$I_1 = \int_0^a f \qquad I_2 = \int_a^\infty f \qquad I_3 = \int_0^a x^2 f$$

$$I_4 = \int_a^\infty x^2 f \qquad I_5 = \int_a^\infty x f$$

Appendix 3.2

Assume for the moment that the truncated range of integration is ignored and (A3.6) is used directly to maximize (3.14) resulting in $b^* < 0$. The effect of ignoring the tail truncation at $\pm x_T$ is equivalent to allowing the nonlinearity g_v to have virtual tails like those illustrated in Fig. 3.7. Additional nonzero tails in g_v are an artifact of the improper range of integration, and as a result, (A3.6) optimizes g_v instead of g_{2l}. This affects the final result b^*, since the detector using g_v would not perform as well as one using g_{2l}, the truncated version. A simple argument explains why: Due to the additional tail area, $Eg'_v \leq Eg'_{2l}$, and $Eg_v^2 \geq Eg_{2l}^2$. Combining these two facts, we find

$$\left[\int_{-\infty}^{\infty} g_v f\right]\left[\int_{-\infty}^{\infty} g_v^2 f\right]^{-1/2} \leq \left[\int_{-\infty}^{\infty} g_{2l} f\right]\left[\int_{-\infty}^{\infty} g_{2l}^2 f\right]^{-1/2} \tag{A3.7}$$

For fixed and equal false alarm rates, g_v will have lower power of detection than g_{2l} asymptotically as the number of samples grows large [14, p. 228]. A weak sufficient condition for the inequality (A3.7) to hold under squaring is

$$b \geq -\int_0^a f \Big/ \int_a^{\infty} f \tag{A3.8}$$

This lower bound on b is b_0 from Appendix I. Since $b^* \in \{b_0, b_1\}$ and $b_1 \geq b_0$, the condition (A3.8) is satisfied for $C < 0$ and $\eta(g_T) \geq \eta(g_v)$. For $C > 0$ it should be possible to prove the observation that $b_0 \leq b_1$, making the squared inequality true for this case also.

A rigorous proof of the inequality (A3.8) under squaring may be found

by considering the effect of the virtual tails in the denominator terms of (A3.7) in addition to their effect in the numerator terms only, as was done here.

References

[1] E.J. Wegman, "Nonparametric Probability Density Estimation: A Survey of Available Methods", *Technometrics*, vol. 14, no. 3, pp. 533-546, August 1972.

[2] M.J. Fryer, "A Review of some Non-Parametric Methods of Density Estimation", *J. Inst. Math. Appl.*, vol. 20, pp. 335-354, 1977.

[3] R.A. Tapia and J.R. Thompson, *Nonparametric Probability Density Estimation*, Baltimore, MD: Johns Hopkins Univ. Press, 1978.

[4] G.R. Wilson and D.R. Powell, "Experimental and Modeled Density Estimates of Underwater Acoustic Returns", University of Texas at Austin, Applied Research Laboratories Tech. Report ARL-TP-82-38, August 1982.

[5] S.L. Bernstein, *et al.*, "Long Range Communications at Extremely Low Frequencies", *Proc. IEEE*, vol. 62, no. 3, pp. 292-312, March 1974.

[6] D. Middleton, "Statistical-Physical Models of Electromagnetic Interference", *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106-127, August 1977.

[7] D. Middleton, "Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance", *IEEE Trans. Electromagn. Compat.*, vol. EMC-21, no. 3, pp. 209-220, August 1979.

[8] M. Abramowitz and I. Stegun, Ed., *Handbook of Mathematical Functions*, New York, NY: Dover Publications, Inc., 1972.

[9] J.H. Miller and J.B. Thomas, "Robust Detectors for Signals in Non-Gaussian Noise", *IEEE Trans. Commun.*, vol. COM-25, no. 7, pp. 686-690, July 1977.

[10] J.W. Modestino, "Adaptive Detection of Signals in Impulsive Noise Environments," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 1022-1026, Sept. 1977.

[11] R.F. Ingram and R. Houle, "Performance of the Optimum and Several Suboptimum Receivers for Threshold Detection of Known Signals in Additive, White, Non-Gaussian Noise," Technical Report 6339, Naval Underwater Systems Center, New London, CT, Nov. 24, 1980.

[12] R.E. Ziemer and R.B. Fluchel, "Selection of Blanking and Limiting Levels for Binary Signaling in Gaussian plus Impulsive Noise", *Proc. IEEE Fall Electronics Conference*, Chicago, IL, pp. 290-295, Oct. 1971.

[13] J. Capon, "On the Asymptotic Relative Efficiency of Locally Optimum Detectors", *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67-71, April 1961.

[14] C.W. Helstrom, *Statistical Theory of Signal Detection*, Oxford; Pergamon Press, 1968.

[15] S.V. Czarnecki and K.S. Vastola, "Approximation of Locally Optimum Detector Nonlinearities", *Proc. 1983 Conf. Inform. Systems and Sciences*, Johns Hopkins Univ, Baltimore MD, pp. 218-221, March 1983.

[16] D.F. Andrews, *et al.*, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, NJ, 1972.

[17] E. J. Modugno III, "The Detection of Signals in Impulsive Noise", *Ph.D. Dissertation*, Dept. of EECS, Princeton University, Princeton NJ, 1982.

[18] K.S. Vastola, "Threshold Detection in Narrowband Non-Gaussian Noise", Tech. Report no. 46, Dept. of EECS, Princeton University, Princeton NJ, 1983. (To appear in *IEEE Trans. Commun.*)

[19] K.S. Vastola, "On Narrowband Impulsive Noise", *Proc. 20th Annual Allerton Conference on Communications, Control, and Computing*, Oct. 1982.

[20] E.J.G. Pitman, *Some Basic Theory for Statistical Inference*, London: Chapman and Hall, 1979.

[21] A. D. Watt and E. L. Maxwell, "Measured Statistical Characteristics of VLF Atmospheric Noise," *Proc. Inst. Radio Engineers*, vol. 45, no. 1, pp 55-62, Jan. 1957.

[22] R. F. Dwyer, "Arctic Ambient Noise Statistical Measurement Results and Their Implications to Sonar Performance Improvements," Technical Report 6739, Naval Underwater Systems Center, New London, CT, May 5, 1982.

4

Signal Detection in Bursts of Impulsive Noise

The previous chapter was concerned with the development of a fixed detector nonlinearity which could adapt to the noise statistics of the particular environment. There it was assumed that, over short periods of time, the statistics were nearly stationary. Another approach to the detection problem is given in this chapter, where it is assumed that the noise statistics can change abruptly.

The fundamental idea explored is that, if the abrupt changes in the noise can be recognized, a detector may use this knowledge to achieve improved performance with respect to a detector whose structure is based upon an assumption of nearly stationary noise statistics. Section 1 provides the background and motivation of the problem. Section 2 develops a model for noise with abruptly changing statistics; specifically, the case of a Gaussian background noise interrupted by bursts of an

impulsive contaminant noise is considered. A detector structure is proposed, and its performance is analyzed. Section 3 examines the problem of distinguishing between impulsive bursts and the background noise. The proposed detector and impulsive burst recognition algorithm are simulated in Section 4, and a few concluding comments are given in Section 5.

1. Introduction

Considerable attention has been paid to the problem of recognizing sudden changes in the stochastic environment of a system. Basseville [1] and Willsky [2] summarize some of the techniques which have been developed. One approach in treating this problem involves the use of characterizations that allow for abrupt changes in the noise statistics. In some simple cases, the noise model consists of two distinct density functions, each describing a unique mode of noise generation. During nonoverlapping time intervals, one of the pair is considered to be the particular valid description of the noise density.

Fig. 4.1 illustrates a conceptual representation of this situation. Only the sequence $\{n_i\}$ may be observed. The sequence $\{e_i\}$ chooses between $n_{0;i}$ and $n_{1;i}$ on a sample-by-sample basis. While $\{e_i\}$ cannot be observed directly, it may be possible to construct an estimate of it by observing the behavior of $\{n_i\}$. A usual assumption is that e_i does not switch "too rapidly"; loosely speaking, after switching into a new state, e_i tends to stay there for a while. It is this property that allows an observer to distinguish between the two noise modes. This assumption is clarified further in Section 3.

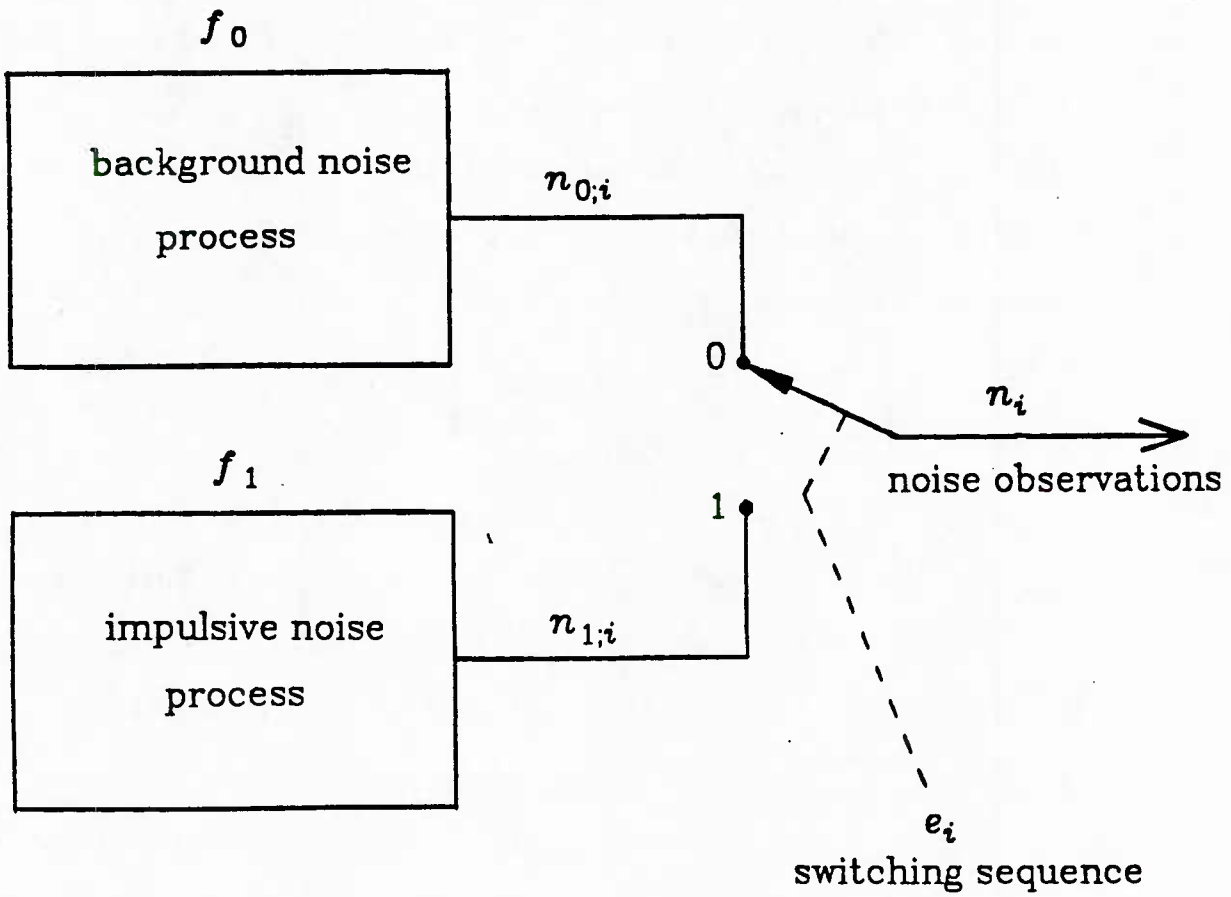

Fig. 4.1. A representation of the dual mode noise generation mechanism.

For a physical example of a noise with abruptly changing statistics, consider the case of the noise environment under the Arctic ice pack [3]: For most of the time, the noise appears to be Gaussian. Occasionally, though, ice cracking occurs, and a short burst of a relatively high variance noise is observed. After the cracking event is complete, the noise returns to a nominal low variance Gaussian mode. Fig. 4.2 reveals the distinctive difference in the behaviors of the two noise modes. This type of noise may be described as an impulsively contaminated Gaussian noise.

Various statistical models have been proposed for describing a noise environment that is nominally Gaussian with an additive impulsive noise component. As was discussed in Chapter 2, these models often take the form of univariate pdf's that are heavy-tailed relative to the Gaussian pdf [*e.g.*, 10,11,16,17]. Implicit with the use of a univariate noise model, however, is the assumption that the noise statistics are stationary at least over the interval of interest. The Arctic under-ice noise is a counterexample to this assumption, since over the short term the noise statistics appear to be nonstationary. The impulsive noise occurs in bursts, and a nonstationary model for the noise seems more appropriate than some fixed model.

It may be possible to find a multivariate noise distribution which adequately describes a background Gaussian noise with bursts of an impulsive contaminant. Unfortunately, finding multivariate non-Gaussian noise models is in general a complicated problem, even in fairly straightforward situations. See, for example, [12-15]. Furthermore, complicated

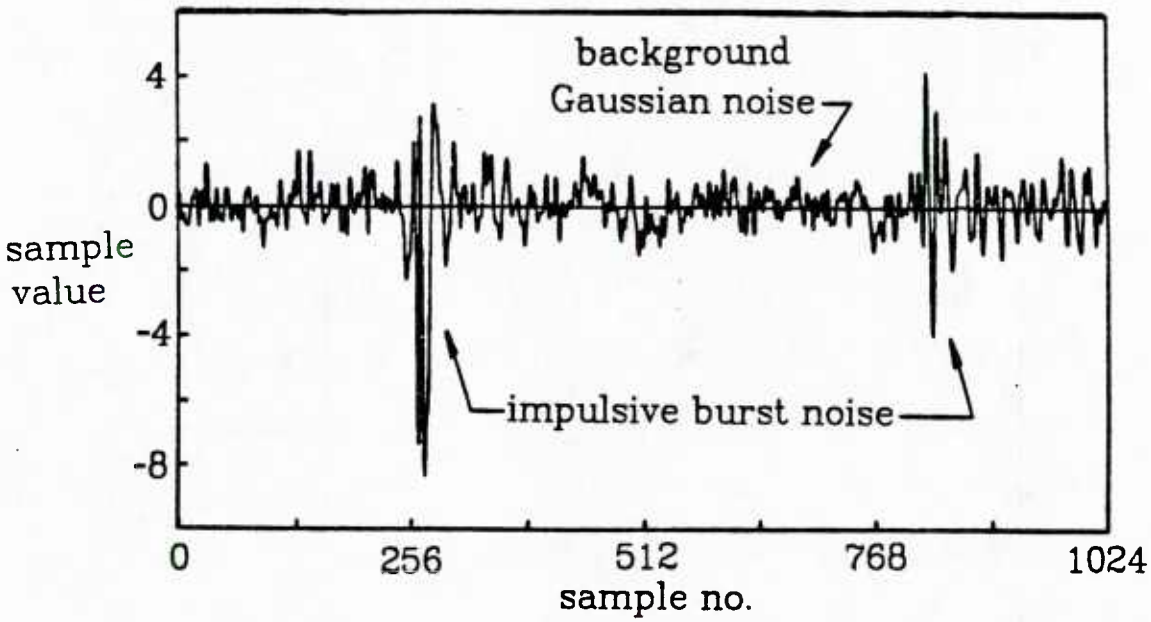

Fig. 4.2. Time domain plot of sample Arctic under-ice noise data record 2220. Vertical scale is in standard deviations from the mean.

multivariate noise distributions may lead to unacceptably complicated optimal detector structures.

When a heavy tailed noise has a burst-like structure as in Fig. 4.2, it is reasonable to develop a detector that recognizes the dual-mode nature of the noise and adapts rapidly to the particular operative mode. The purpose of this chapter is to illustrate the potential advantage of such a *switched burst* (*SB*) detector.

2. Switched Burst Detector

We shall restrict attention to the discrete time locally optimal (LO) detection of a known constant signal in a Gaussian background noise contaminated by bursts of impulsive noise. All the noise samples are assumed independent, but not necessarily identically distributed. In more precise form, the problem is to observe $\mathbf{x}=x_i,\ i=1,2,\ldots,M$ and decide between

$$H_0\colon \mathbf{x}=\mathbf{n}$$
$$H_1\colon \mathbf{x}=\mathbf{n}+\theta\mathbf{s}$$

where $\mathbf{n}=n_i,\ i=1,2,\ldots,M$ and $\mathbf{s}$ is a known constant signal s of length n and nonzero amplitude parameter θ. As is well known [5], the LO detector test statistic in the case of white noise is any monotone function of

$$T_n = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln \frac{f_i(x_i-\theta s)}{f_i(x_i)}\bigg|_{\theta=0}$$

where f_i is the univariate density of n_i. The term being summed is a memoryless nonlinearity $g_i(x_i) = -\dfrac{f_i'(x_i)}{f_i(x_i)}$.

Stationary Impulsive Model

One common empirical model of impulsively contaminated Gaussian noise is the Gaussian-Gaussian ε-mixture density, which may be written as

$$f_{\varepsilon}(x) = (1-\varepsilon)f_0(x)+\varepsilon f_1(x) \tag{4.1}$$

Here f_0 represents a background low variance Gaussian noise, and f_1 represents a high variance (impulsive) Gaussian component. Any particular observation x is generated by the impulsive component with probability ε. Vastola [4] recently suggested that (4.1) is also a useful simplification of Middleton's Class A model [10,11] for impulsive noise environments.

Using (4.1) as the univariate density of the noise, and assuming that the noise samples are identically distributed, the LO detector nonlinearity is fixed for all samples as

$$g_{\varepsilon}(x) = x\left[\frac{\frac{(1-\varepsilon)}{\sigma_0^2}f_0(x)+\frac{\varepsilon}{\sigma_1^2}f_1(x)}{(1-\varepsilon)f_0(x)+\varepsilon f_1(x)}\right] \tag{4.2}$$

For convenience, the overall noise variance is assumed to be unity.

Switched Burst Nonstationary Model

Consider the nonstationary noise density

$$f_{SB}(x_i;e_i) = (1-e_i)f_0(x_i)+e_i f_1(x_i) \tag{4.3}$$

for $i = 1,2,.....$ Here, e_i takes on the value 0 or 1, and f_0 and f_1 are two arbitrary densities, which are not necessarily Gaussian. When e_i is zero,

the noise is in the *background mode*, and the observed noise has density f_0. When e_i is unity, the noise is in the *impulsive mode* and the observed noise density becomes f_1. The sequence $\{e_i\}$ is defined to have the property

$$\varepsilon = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} e_i \tag{4.4}$$

The implication of (4.4) is that, over a long observation period, ε proportion of the samples come from the impulsive noise mode, and $(1-\varepsilon)$ proportion of the samples come from the background mode. Noise samples described by f_{SB} may be thought of as being generated by the mechanism of Fig. 4.1. In many cases of interest, the background mode is dominant, and therefore, ε is often estimated or assumed to be small [4,8,17]. Note that unlike f_ε, the density f_{SB} is nonstationary. However, the noise in each individual mode may be considered stationary, with the sequence $\{e_i\}$ controlling which mode is observed.

For the purpose of comparison with the Gaussian-Gaussian ε-mixture density, f_0 and f_1 shall be the same densities as those composing f_ε. One rationale for picking f_0 and f_1 to be Gaussian is that they are the two leading and most significant terms in Middleton's Class A density model [4,24]. Equivalently, this assumption implies that the impulsive contaminant is itself a Gaussian noise source. The case of f_0 and f_1 both Gaussian shall be designated as *Gaussian-Gaussian switched burst noise*.

The observations x_i will continue to be assumed independent for any arbitrary switching sequence $\{e_i\}$. The noise density on a sample by sample basis is either f_0 or f_1, which requires that the nonlinearity used at a

particular sample time is g_0 or g_1, respectively. Ideally, g_0 and g_1 are the locally optimal nonlinearities associated with the two densities. The test statistic becomes

$$T_n = \sum_{i=1}^{n} g_{SB}(x_i; e_i) \tag{4.5}$$

where

$$g_{SB} = \begin{cases} g_0 & \text{for } e_i = 0 \\ g_1 & \text{for } e_i = 1 \end{cases} \tag{4.6}$$

In practice, $\{e_i\}$ would not be known; instead, some additional structure is required to generate a sequence $\{p_i\}$ as an estimate of $\{e_i\}$. This problem receives attention in Section 3.

Ideal Detector Performance

In this subsection, expressions are given for the performance of the switched burst detector with the assumption that the switching sequence may be reconstructed without error. Performance will be analyzed for arbitrary densities f_0 and f_1 with arbitrary nonlinearities g_0 and g_1. Specific results for Gaussian-Gaussian switched burst noise with linear detectors will also be developed. The next subsection explores detector performance without this ideal knowledge.

A definition for the efficacy of an arbitrary stationary detector g with zero mean under the noise f given in [5,6], and discussed in Chapter 2 is

$$\eta_f(g) = \frac{\mathrm{E}_f^2 g'}{\mathrm{E}_f g^2} \tag{4.7}$$

The case of interest is nonstationary, as the nonlinearities composing g_{SB}

switch in accordance with the switching of the underlying noise densities. Appendix 6.1 demonstrates that, for the switching detector in the presence switched burst noise, (4.7) may be rewritten as

$$\eta_{SB}(g_{SB}) = \frac{\left[(1-\varepsilon)E_{f_0}g_0' + \varepsilon E_{f_1}g'_1\right]^2}{(1-\varepsilon)E_{f_0}g_0^2 + \varepsilon E_{f_1}g_1^2} \tag{4.8}$$

The formulation (4.8) is general, and does not depend on the fact that f_0 and f_1 were previously defined to be Gaussian densities. For the particular case of the Gaussian densities used in (4.1) and (4.3), the LO detector is linear with slope σ_0^{-2} for f_0 and with slope σ_1^{-2} for f_1. Applying this fact to (4.8), the expression for η_{SB} reduces to

$$\eta_{SB} = \left[\frac{1-\varepsilon}{\sigma_0^2} + \frac{\varepsilon}{\sigma_1^2}\right] \tag{4.9}$$

Note that (4.8) may also be used to evaluate the performance of g_ε in the presence of f_{SB}. In this case, the detector nonlinearity is the same whether the background or impulsive noise is observed; therefore we can let $g_0 = g_1 = g_\varepsilon$ in (4.8).

A convenient measure for comparing the performance of two LO detectors d_1 and d_2 is asymptotic relative efficiency, defined in Chapter 2 as

$$\mathrm{ARE}_{d_1,d_2} = \eta_{d_1} / \eta_{d_2} \tag{4.10}$$

The two nonlinearities g_ε and g_{SB} may be compared by computing their efficacies and evaluating (4.10).

Fig. 4.3 presents a plot of $\mathrm{ARE}_{g_{SB},g_\varepsilon}$ for combinations of ε and σ_1^2,

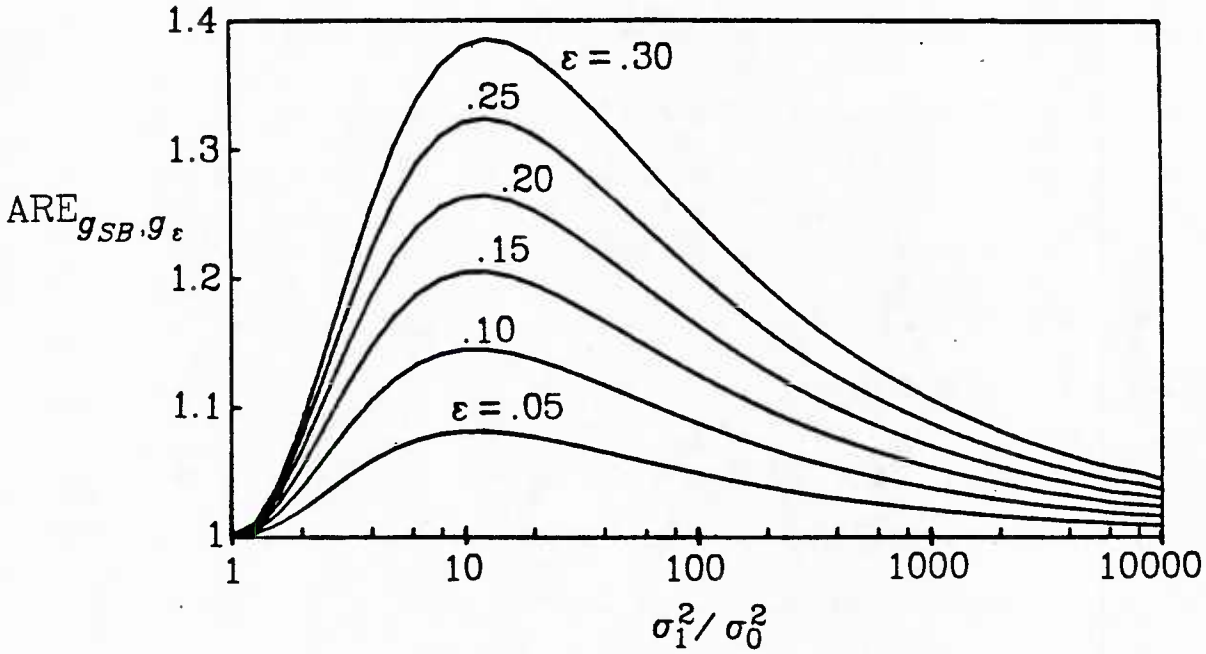

Fig. 4.3. *Performance comparison of fixed nonlinearity* g_ε *and switched nonlinearity* g_{SB} *in Gaussian-Gaussian switched burst noise for various values of* ε *and range of* σ_1^2/σ_0^2.

from which it is clear that there is an advantage to the switched detector method over the fixed detector. An intuitive explanation for this is that in switched burst noise, the noise is always Gaussian (though with nonstationary variance), and the LO detector is always linear (with nonstationary slope). Thus, the switched detector maximizes efficacy for it is always locally optimal at any given sample time. On the other hand, g_ε has two nearly linear regions, but it is not the locally optimal detector for Gaussian noise. If the stationary density f_ε is thought of as a time averaged version of f_{SB}, then in some sense the nonlinearity g_ε may be interpreted as an optimal stationary approximation of g_{SB}. The three nonlinearities g_0, g_1, and g_ε are plotted in Fig. 4.4.

Another point not made obvious by Fig. 4.3 is that the switched detector is capable of large performance improvements over a fixed linear detector, *ld*. A plot of $\mathrm{ARE}_{g_{SB},ld}$ is presented in Fig. 4.5. For very large values of σ_1^2/σ_0^2, the switched detector has a processing gain relative to *ld* of approximately $\varepsilon\sigma_1^2/\sigma_0^2$.

Non-Ideal Detector Performance

The previous subsection discussed the performance of the switched detector under the ideal assumption of perfect knowledge of $\{e_i\}$, and that g_0 and g_1 were LO for f_0 and f_1, respectively. In any practical situation, these assumptions would almost certainly be violated. As a result, we now direct attention towards the performance of g_{SB} when only an estimate $\{p_i\}$ of $\{e_i\}$ is available. Also, the effects of incorrectly estimating the variance ratio in the Gaussian-Gaussian switched burst noise is

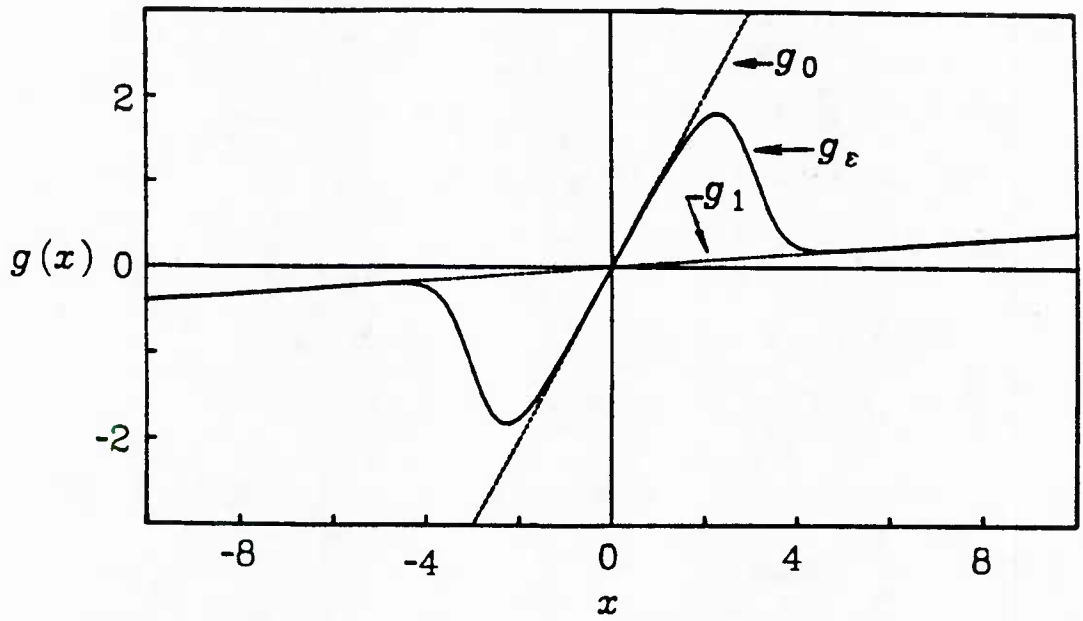

Fig. 4.4. The nonlinearity g_ε compared to the two linear detectors g_0 and g_1. The slopes of g_0 and g_1 are σ_0^{-2} and σ_1^{-2}, respectively.

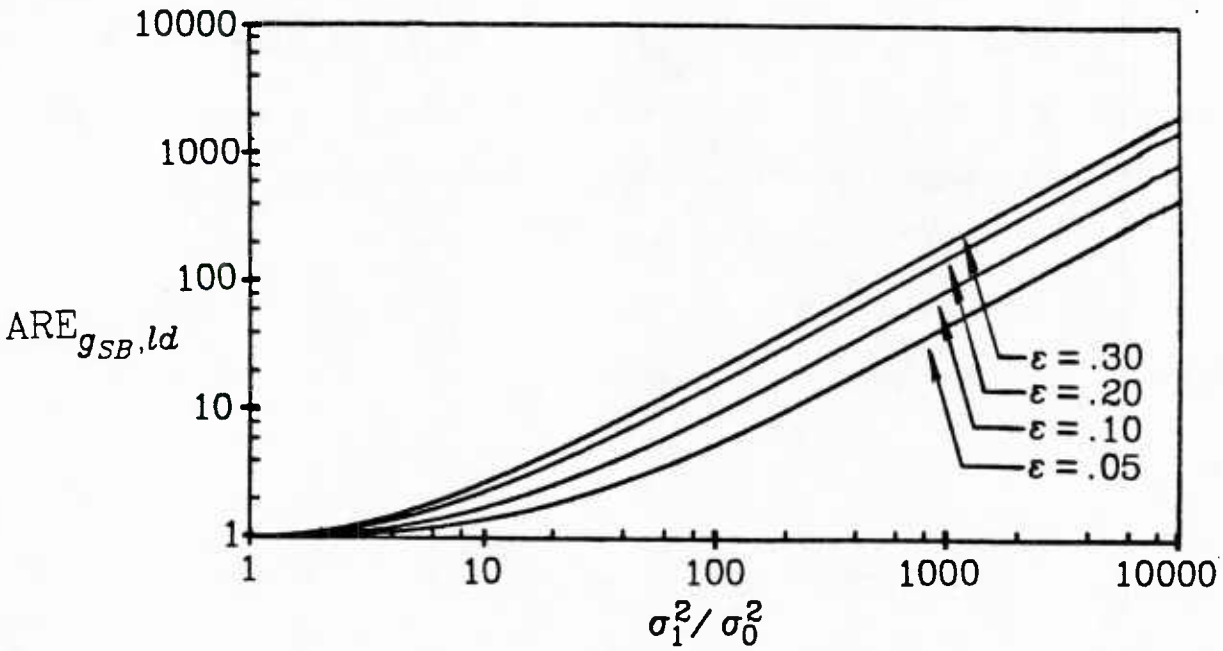

Fig. 4.5. Comparison of the switched detector g_{SB} and the linear detector *ld* in Gaussian-Gaussian switched burst noise for various values of ε and range of σ_1^2/σ_0^2.

assessed.

Appendix 6.1 gives a detailed development of η_{SB} assuming $\{e_i\}$ is known. Appendix 6.2 notes the changes in the development arising when the estimated sequence $\{p_i\}$ is used, and obtains the result

$$\eta_{SB} = \frac{\left[(1-\varepsilon)E_{f_0}\left[(1-\rho_{1|0})g_0'+\rho_{1|0}g_1'\right] + \varepsilon E_{f_1}\left[\rho_{0|1}g_0'+(1-\rho_{1|0})g_1'\right]\right]^2}{(1-\varepsilon)E_{f_0}\left[(1-\rho_{1|0})g_0^2+\rho_{1|0}g_1^2\right] + \varepsilon E_{f_1}\left[\rho_{0|1}g_0^2+(1-\rho_{0|1})g_1^2\right]} \tag{4.11}$$

where $\rho_{1|0}$ is the probability of using nonlinearity g_1 when the true noise density is f_0, and $\rho_{0|1}$ is the probability of using nonlinearity g_0 when the true noise density is f_1. When both error probabilities are identically zero, then (4.11) reduces to the special case (4.8) of operation without switching error.

Paralleling the discussion of the previous subsection, we shall continue to use the Gaussian-Gaussian switched burst model. Then f_0 and f_1 are Gaussian densities, and g_0 and g_1 again are linear detectors with slopes σ_0^{-2} and σ_1^{-2}, respectively. The situation may be generalized slightly by assuming that g_1 has slope $\hat{\sigma}_1^2$ not necessarily equal to σ_1^2. No additional generality in the efficacy calculation results by allowing the slope of g_0 to vary from σ_0^{-2}, as only the ratio of slopes affects performance. Under these assumptions, (4.9) generalizes to

$$\eta_{SB} = \frac{\left[(1-\varepsilon)\left(\frac{1-\rho_{1|0}}{\sigma_0^2}+\frac{\rho_{1|0}}{\hat{\sigma}_1^2}\right)+\varepsilon\left(\frac{\rho_{0|1}}{\sigma_0^2}+\frac{1-\rho_{0|1}}{\hat{\sigma}_1^2}\right)\right]^2}{(1-\varepsilon)\left(\frac{1-\rho_{1|0}}{\sigma_0^2}+\frac{\rho_{1|0}\sigma_0^2}{\hat{\sigma}_1^4}\right)+\varepsilon\left(\frac{\rho_{0|1}\sigma_1^2}{\sigma_0^4}+\frac{(1-\rho_{0|1})\sigma_1^2}{\hat{\sigma}_1^4}\right)} \tag{4.12}$$

For the purposes of comparison, SB_e will denote the the switched burst

detector with error, and SB_i will denote the ideal error-free switched burst detector. The noise environment will be the Gaussian-Gaussian switched burst environment, with $\varepsilon = 0.1$ chosen as a value giving representative numerical results.

To begin, the effect of the errors $\rho_{0|1}$ and $\rho_{1|0}$ will be examined. Fig. 4.6 presents a plot of ARE_{SB_e, SB_i} for a range of values of $\rho_{0|1}$, with $\rho_{1|0}$ fixed at zero, and $\hat{\sigma}_1^2 = \sigma_1^2$. Here, the effect of incorrectly choosing to use non-linearity g_0 during the impulsive mode is isolated. Two conclusions are obvious: first, performance deteriorates monotonically with increasing error probability $\rho_{0|1}$. Second, the effect of not recognizing an impulsive noise observation is much worse as the variance ratio increases. However, it is reasonable to assume that as the variance ratio increases, it is easier for an algorithm to recognize noise bursts, and therefore $\rho_{0|1}$ will be small.

Figure 4.7 demonstrates the effects of deciding incorrectly that a noise burst is present. Here, $\rho_{1|0}$ is allowed to vary while $\rho_{0|1} = 0$, and $\hat{\sigma}_1^2 = \sigma_1^2$. Clearly, making this type of error is far less damaging to performance than deciding incorrectly that the noise is in background mode.

The combined effects of the two errors may be seen in Fig. 4.8. Here, $\rho_{0|1} = 0.02$, and $\rho_{1|0}$ varies. The results are consistent with the results of Figs. 4.6 and 4.7: the performance deterioration is due mainly to $\rho_{0|1}$, with a $\rho_{1|0}$ providing a lesser deterioration. The conclusion which may be drawn from these three figures is that correctly recognizing the presence of an impulsive burst is of critical importance to the success of the switched burst detector. Note that this conclusion gives support to the

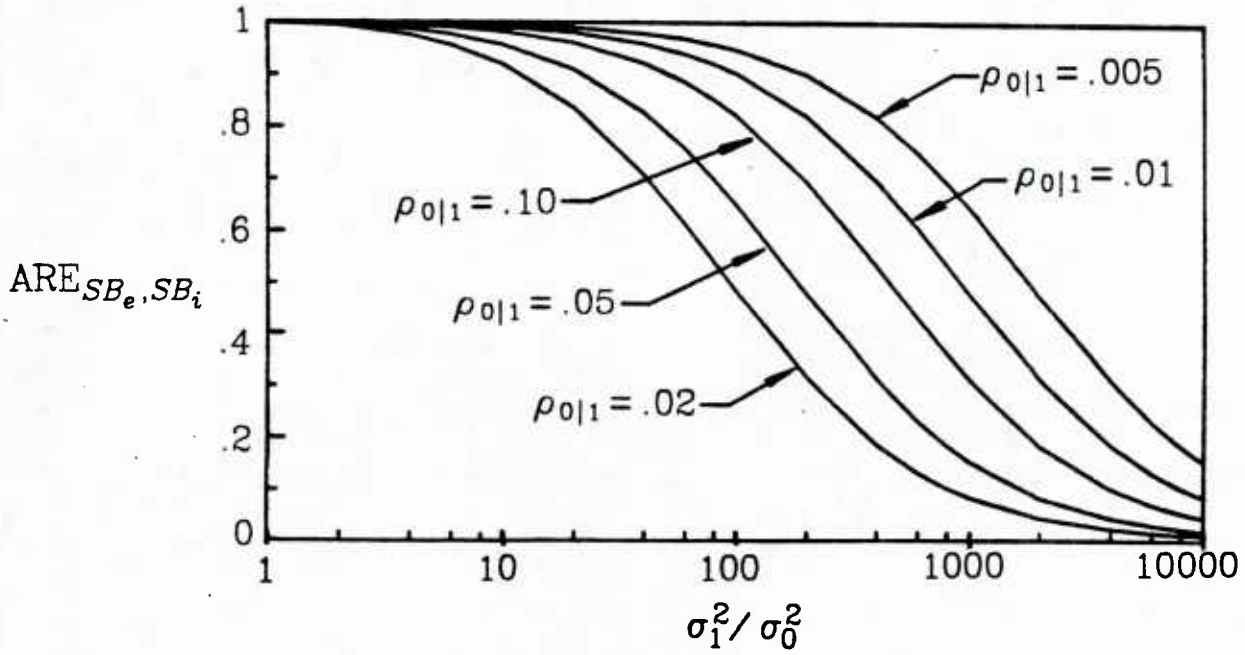

Fig. 4.6 Performance of switched detector with errors SB_e relative to ideal switched detector SB_i for various probabilities $\rho_{0|1}$ of incorrectly classifying an impulsive noise sample as background noise.

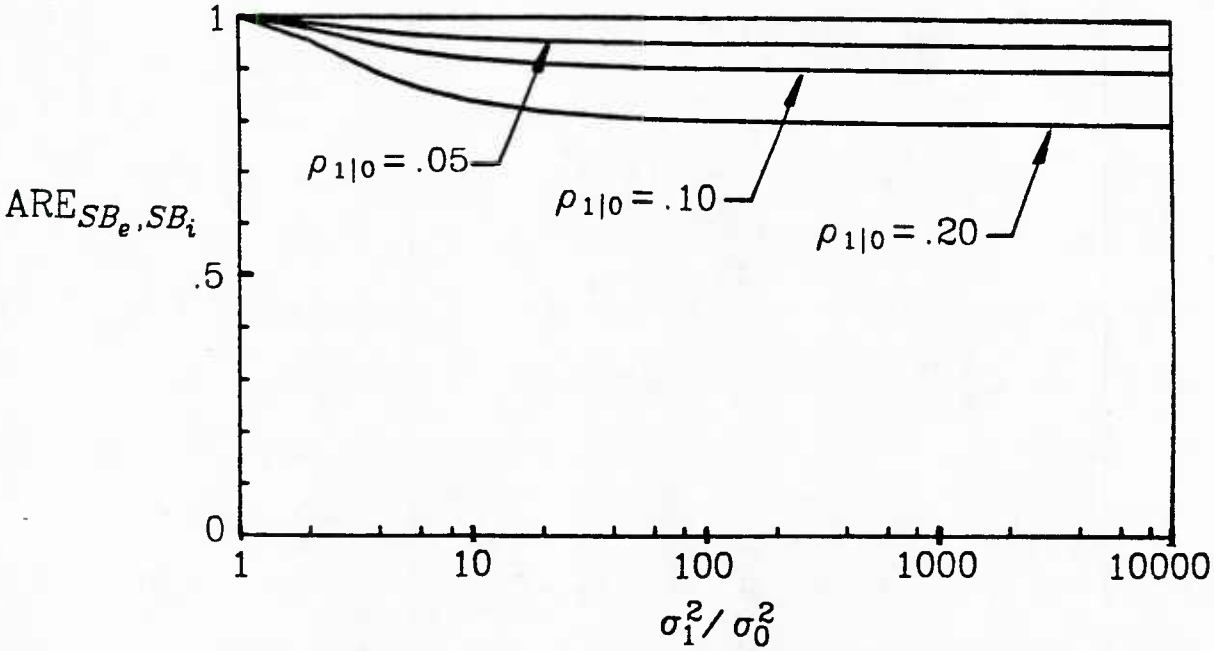

Fig. 4.7. Performance of switched detector with errors SB_e relative to ideal switched detector SB_i for various probabilities $\rho_{1|0}$ of incorrectly classifying a background noise sample as impulsive.

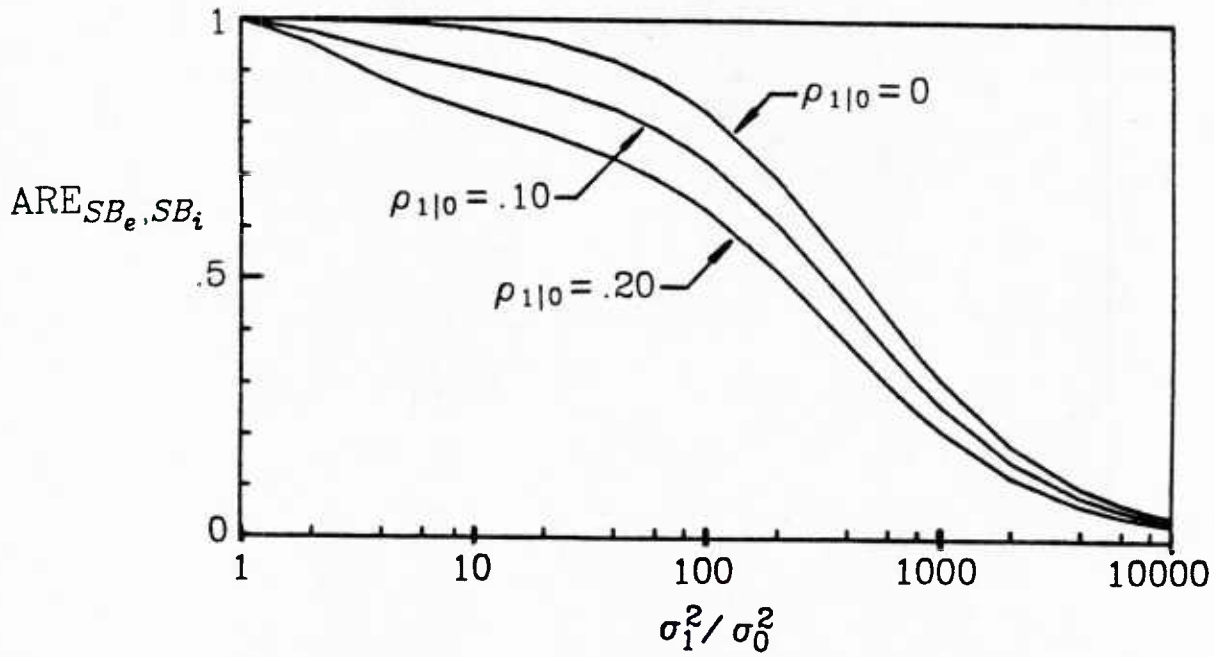

Fig. 4.8. Performance of switched detector with errors SB_e relative to ideal switched detector SB_i for various probabilities $\rho_{0|1}$ of incorrectly classifying a background noise sample as impulsive with fixed probability $\rho_{0|1} = .02$ of classifying an impulsive noise sample as background noise.

intuitive notion that even a few impulsive noise samples can seriously disturb detector performance.

The effect of incorrectly choosing the slope of g_1 will now be examined. Here, $\rho_{0|1} = \rho_{1|0} = 0$, and $\hat{\sigma}_1^2 / \sigma_1^2$ is allowed to vary. Fig. 4.9 gives ARE_{SB_e, SB_i} for small values of the variance ratio, and Fig. 4.10 presents the case of large variance ratios. As illustrated, $\hat{\sigma}_1^2 / \sigma_1^2$ may deviate significantly from unity, with only moderate effects on performance, especially for instances where $\hat{\sigma}_1^2 \gg \sigma_1^2$. Note that, when the ratio approaches infinity, g_{SB} essentially "turns off" during impulsive bursts. Further, as σ_1^2 / σ_0^2 grows large, the effect of incorrect $\hat{\sigma}_1^2$ diminishes. Surprisingly, estimating $\hat{\sigma}_1^2$ inaccurately does not critically affect performance. An implication of Fig. 4.9 and 4.10 is that, when $\hat{\sigma}_1^2$ must be estimated from the noise data, good asymptotic detector performance may be maintained simply by biasing the estimate towards large values.

3. Discrimination between Noise Modes

In this section, an algorithm will be developed to regenerate the switching sequence $\{e_i\}$. It was previously assumed that the sequence $\{e_i\}$ did not switch "too rapidly". The assumption may be interpreted here as meaning that the probability of a very short run of ones in $\{e_i\}$ is negligible.

Parametric Modeling

There are a number of ways to model the statistics of sequence $\{e_i\}$. Gilbert [19] proposes a Markov chain taking on one of the two state values $\{Background, Impulsive\}$, corresponding to the proposed states zero and

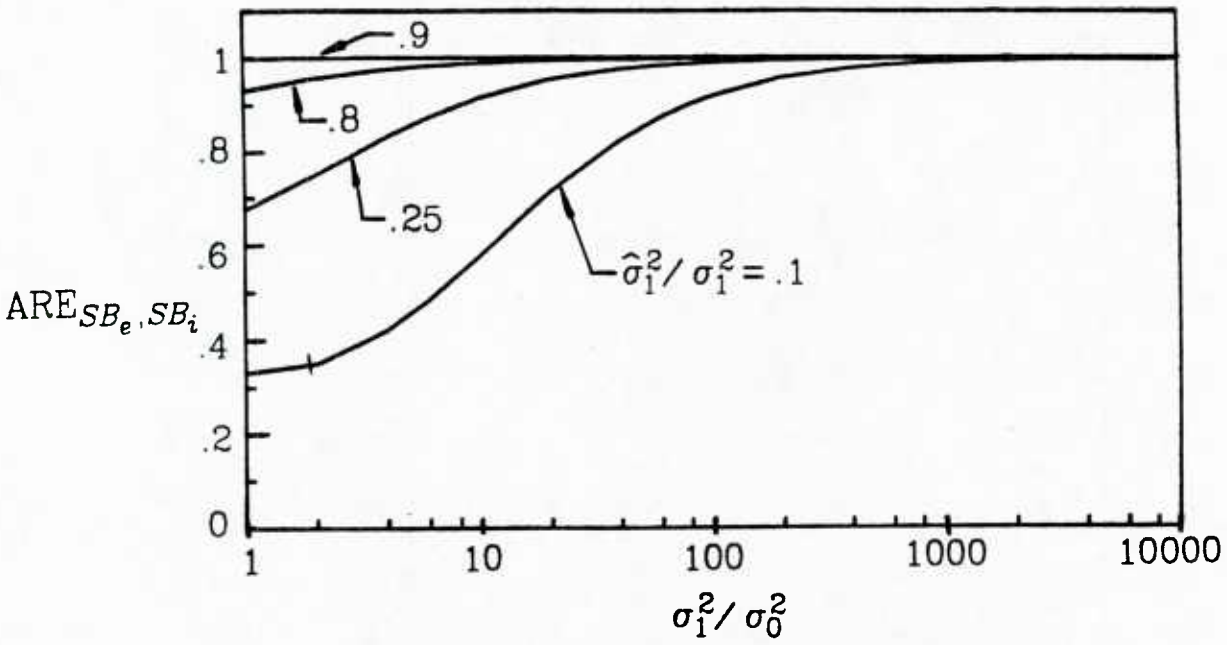

Fig. 4.9. Performance of switched detector with errors SB_e relative to ideal switched detector SB_i for various errors $\hat{\sigma}_1^2/\sigma_1^2 < 1$ of impulsive variance estimate.

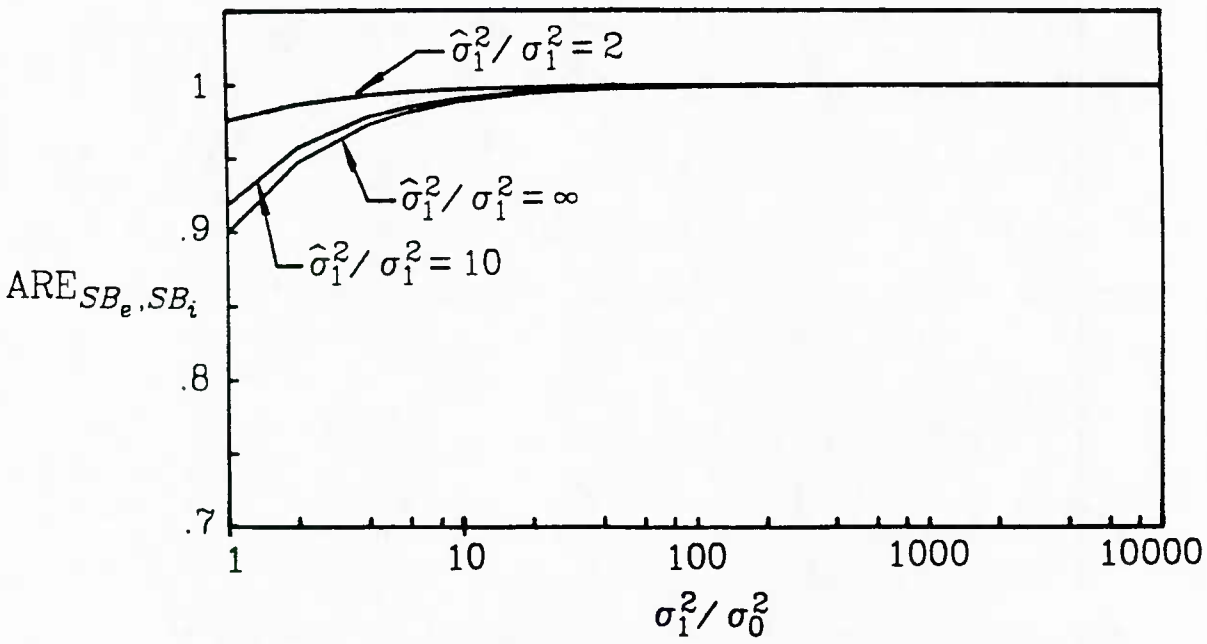

Fig. 4.10. Performance of switched detector with errors SB_e relative to ideal switched detector SB_i for various errors $\hat{\sigma}_1^2/\sigma_1^2 > 1$ of impulsive variance estimate.

one, respectively. This model was used with success by Ehrman [20] in simulating an impulsively contaminated Gaussian channel. During a run of a particular state in $\{e_i\}$, each observation e_i may be considered as the outcome of a Bernoulli trial, with probability of success equal to one of the two state transition probabilities $p_{1\to0}$ or $p_{0\to1}$. As is well known [21], under this condition it follows that the run length, or residence time, of each state has a geometric probability density. The geometric density itself is a particular case of the negative binomial density. If the transition probabilities $p_{1\to0}$ and $p_{0\to1}$ are different, the residence time density for each state has a different negative binomial density.

Another natural model for the run length statistics of $\{e_i\}$ is the Poisson density, where the rate parameter of the density is the mean state residence time. The rate parameters of the two states need not be identical. If the rate parameter of a Poisson density is not known exactly, and instead is distributed as a gamma density, then the compound density is the negative binomial density [21, pp. 122-3]; if the rate parameter is exponentially distributed (a special case of the gamma density), then the compound density is the geometric density. As a result, the negative binomial density is sometimes referred to as a gamma-mixture Poisson density.

For equal means, the variance of the geometric density is greater than the variance of the Poisson. This is to be expected, as use of the Poisson density model implies perfect knowledge of the state mean run length, while the geometric model implies that only a statistical description of the state mean run length is available.

Using parametric statistical models for $\{e_i\}$, tests may be devised which observe a noise sequence $\{n_i\}$, and generate $\{p_i\}$, an estimate of the switching sequence. Various approaches include sample-by-sample tests, pattern recognition approaches, and maximum likelihood sequence reconstruction. With an accurate model, these approaches may quite accurately reconstruct $\{e_i\}$. The difficulty, however, is that fairly detailed information about the statistics of $\{e_i\}$ may be needed, and often this information may be unavailable.

Nonparametric Approach

The advantage to a nonparametric approach is that detailed statistical information is not necessary to construct a test, and that nonparametric tests are usually fairly robust: they work reasonably well over a broad range of situations. Furthermore, they often have simple structures. On the other hand, they are generally less efficient than optimal decision structures; *i.e*, given the same amount of data, there is a higher probability of error.

The following two-step algorithm is proposed to generate $\{p_i\}$, a reconstruction of $\{e_i\}$:

(*i*) On a sample-by-sample basis, decide between

$$n_i \sim f_0 \quad \text{say } \hat{p}_i = 0 \qquad (4.13)$$

$$n_i \sim f_1 \quad \text{say } \hat{p}_i = 1$$

(*ii*) Filter the sequence $\{\hat{p}_i\}$ to obtain $\{p_i\}$. Here, *filter* means to smooth in a manner which tends to reduce the

number of incorrect state transitions.

If f_0 and f_1 are known, then (4.13) may be carried out by a likelihood ratio test. In the case where f_0 and f_1 are both Gaussian, and $\sigma_0^2 \neq \sigma_1^2$, the test becomes

$$|n_i| \underset{\hat{p}_i=0}{\overset{\hat{p}_i=1}{\gtrless}} T \tag{4.14}$$

To filter $\{\hat{p}_i\}$, perform the test

$$\sum_{k=-m}^{m} \hat{p}_{i+k} \underset{p_i=0}{\overset{p_i=1}{\gtrless}} m \tag{4.15}$$

for some integer $m \geq 0$. If $m = 0$, no smoothing occurs, and $\hat{p}_i = p_i$ for every i. If $m > 0$, the test (4.14) is a *voting* algorithm, where the outcome is the majority state in a window of length $2m+1$, centered about $\hat{p}_i$. Since $\hat{p}_i \in \{0,1\}$, the smoothing algorithm is a special case of the *median filtering* algorithm; properties of median filtering have been studied recently by Gallagher and Wise [23]. It will tend to preserve transitions into a new state with run length greater than $m+1$ (edge-preserving property), while tending to suppress runs with length less than m (impulse filtering property).

Nonparametric Algorithm Analysis

To calculate the error statistics of the smoothed sequence $\{p_i\}$, it is first necessary to calculate the performance of the sample-by-sample test (4.14). As it is a binary hypothesis test on individual observations elements of $\{n_i\}$, two errors are possible:

$$\hat{\rho}_{1|0} = \text{Prob}(\text{ say } p_i = 1 \mid e_i = 0) = 2[1-F_0(T)] \tag{4.16}$$

$$\hat{\rho}_{0|1} = \text{Prob}(\text{ say } p_i = 0 \mid e_i = 1) = 2F_1(T)-1 \tag{4.17}$$

where F is the cumulative distribution, and the densities are symmetric about zero. The correct decision probabilities are given by

$$\hat{\rho}_{1|1} = 1-\hat{\rho}_{0|1} \tag{4.18}$$

$$\hat{\rho}_{0|0} = 1-\hat{\rho}_{1|0}$$

Note that the error probabilities are a function only of the value of e_i.

Calculating the performance of the filtered sequence is more complex. For convenience, we first define the vectors

$$\mathbf{e}_i = (e_{i-m}, \ldots, e_{i+m})$$

and

$$\hat{\mathbf{p}}_i = (\hat{p}_{i-m}, \ldots, \hat{p}_{i+m})$$

and the length $2m+1$ vectors

$$\mathbf{0} = (0, \ldots, 0)$$

$$\mathbf{1} = (1, \ldots, 1)$$

If $\mathbf{e}_i = \mathbf{0}$, each outcome in $\hat{\mathbf{p}}_i$ is the result of a Bernoulli trial with a constant probability of success on each trial. A particular element in the filtered outcome $\{p_i\}$ will be in error only if at least $m+1$ elements in $\hat{\mathbf{p}}_i$ are unity. The error probability is given by the cumulative binomial distribution with constant probabilities in each trial.

$$\rho_{1|0}(i) = \sum_{k=m+1}^{2m+1} \binom{2m+1}{k} (\hat{\rho}_{1|0})^k (\hat{\rho}_{0|0})^{2m-k+1} \tag{4.19}$$

Similarly, for $\mathbf{e}_i = \mathbf{1}$,

$$\rho_{0|1}(i) = \sum_{k=m+1}^{2m+1} \binom{2m+1}{k} (\hat{\rho}_{0|1})^k (\hat{\rho}_{1|1})^{2m-k+1} \tag{4.20}$$

The two remaining cases are for nonhomogeneous $\mathbf{e}_i$. This situation occurs when the $2m+1$ filter window contains a state transition of $\{e_i\}$; for example, $\mathbf{e}_i = (0, \ldots, 0, 1, \ldots, 1)$.

For the first remaining case, suppose $e_i = 1$, and let $m_0 + m_1 = 2m+1$, where m_0 is the number of zeroes in $\mathbf{e}_i$, and m_1 the number of ones. Assuming that the state run lengths are greater than $2m+1$, a state transition in $\mathbf{e}_i$ means there are m_0 zeroes followed by m_1 ones, or vice versa. The noise observations $\{n_i\}$ are independent and each $\hat{p}_i$ is an independent Bernoulli trial outcome. However, $\mathbf{e}_i$ is not homogeneous, and the probabilities of the outcomes of $\hat{p}_i$ vary; therefore, p_i is distributed as the outcome of a binomial experiment of $2m+1$ Bernoulli trials with variable probabilities of success [22, p. 282]. The statistic $\sum_{k=-m}^{m} \hat{p}_{i+k}$ in the test (4.15) may be thought of as being the sum $p_0 + p_1$, where p_0 is the binomially distributed outcome of m_0 Bernoulli trials with constant probabilities, and p_1 the outcome of m_1 trials. The probability of making an error, given that $e_i = 1$, is

$$\rho_{0|1}(i) = \sum_{j=m+1}^{2m} \sum_{k_0+k_1=j} \binom{m_0}{k_0} (\hat{\rho}_{0|0})^{k_0} (\hat{\rho}_{1|0})^{m_0-k_0} \binom{m_1}{k_1} (\hat{\rho}_{0|1})^{k_1} (\hat{\rho}_{1|1})^{m_1-k_1} \tag{4.21}$$

where $0 \le k_0 \le m_0$ and $0 \le k_1 \le m_1$. The summation indices k_0 and k_1 may be interpreted as the number of times in $\hat{\mathbf{p}}_i$ that $\hat{p}_{i+k} = 0$ given $e_{i+k} = 0$, and the number of times $\hat{p}_{i+k} = 1$ given $e_{i+k} = 1$, respectively, with

$-m \le k \le m$. If $m_0 = 0$, then (4.21) specializes to (4.20), the probability of deciding $p_i = 0$ when $\mathbf{e}_i = \mathbf{1}$.

Treatment of the second remaining case of nonhomogeneous $\mathbf{e}_i$ is similar. Given $e_i = 0$, the probability of deciding $p_i = 1$ is given by

$$\rho_{1|0}(i) = \sum_{j=m+1}^{2m+1} \sum_{k_0+k_1=j} \binom{m_0}{k_0} (\hat{\rho}_{1|0})^{k_0} (\hat{\rho}_{0|0})^{m_0-k_0} \binom{m_1}{k_1} (\hat{\rho}_{1|1})^{k_1} (\hat{\rho}_{0|1})^{m_1-k_1} \qquad (4.22)$$

where the indices k_0 and k_1 may be interpreted here as the number of times in $\hat{\mathbf{p}}_i$ that $\hat{p}_{i+k} = 1$ given $e_{i+k} = 0$, and the number of times $\hat{p}_{i+k} = 0$ given $e_{i+k} = 1$, respectively, for $-m \le k \le m$. If $\hat{\rho}_{0|1} = \hat{\rho}_{1|0}$, then (4.21) and (4.22) are symmetric in m_0 and m_1.

While the error probability $\hat{\rho}$ for elements of the unsmoothed sequence $\{\hat{p}_i\}$ is a function only of e_i, after smoothing the sequence the error probability $\rho(i)$ is a function of the subsequence $\mathbf{e}_i$. The performance analysis of Appendix 4.2 requires time invariant error probabilities. If the statistics of the state run lengths of $\{e_i\}$ are known, then the expectation of $\rho_{0|1}(i)$ and $\rho_{1|0}(i)$ may be taken and used in evaluating (4.12).

The choice of filter length $2m+1$ affects the error performance. There are two competing considerations: on the one hand, it is desirable to make $2m+1$ as large as possible, for the error probabilities decrease with increasing filter length, provided a state transition does not occur within $\mathbf{e}_i$. On the other hand, making $2m+1$ small reduces the probability of making errors in the vicinity of a state transition.

The following argument will assist in choosing $2m+1$: For the test

(4.15) to recognize a state transition, at least $m+1$ elements of $\hat{\mathbf{p}}_i$ must take on the value of the new state. If the last $m+2$ elements of $\mathbf{e}_i$ belong to the new state, on the average the last $(m+2)\hat{\rho}_{x|x}$ elements of $\hat{\mathbf{p}}_i$ will take on the new state value, and conversely if the first $m+2$ elements of $\mathbf{e}_i$ take the value of the old state. It is reasonable to choose m so that $(m+2)\hat{\rho}_{x|x} \geq m+1$, ensuring that $\hat{\mathbf{p}}_{i\pm2}$ contains on the average at least $m+1$ correct state decisions, given that a state transition occurs between e_i and $e_{i\pm1}$.

A simple manipulation shows this condition is equivalent to

$$2m+1 \leq \min_{x\in\{0,1\}} \frac{3\hat{\rho}_{x|x}-1}{1-\hat{\rho}_{x|x}} \tag{4.23}$$

The minimum nontrivial filter length is $2m+1=3$, which leads to the requirement that $\min_{x\in\{0,1\}} \hat{\rho}_{x|x} \geq \frac{2}{3}$.

Performance of the Nonparametric Algorithm

As was shown in the last subsection, three parameters determine the error performance of the sequence estimation algorithm: the sample-by-sample decision errors $\hat{\rho}_{0|1}$ and $\hat{\rho}_{1|0}$, and the filter length $2m+1$. The effect of these parameters is examined in the following set of figures.

The performance of the threshold test (4.14) is a function of the threshold T, and the densities f_0 and f_1. In particular, when the two densities are zero mean Gaussian densities with variance ratio σ_1^2/σ_0^2, the decision probabilities become

$$\hat{\rho}_{1|0} = 2\Phi(-T/\sigma_0) \tag{4.24}$$

and

$$\hat{\rho}_{1|1} = 2\Phi(-T/\sigma_1) \tag{4.25}$$

where Φ is the cumulative distribution function of the unit variance Gaussian density. It is natural to present the performance of the test (4.14) via a set of receiver operating curves, shown in Fig. 4.11. As is intuitively obvious, and clear from the figure, the probability of recognizing an impulsive sample increases as the distinction between background and impulsive variances increases.

The next three figures consider the effects of $\hat{\rho}_{0|1}$ and $\hat{\rho}_{1|0}$ upon the performance of the filtered sequence $\{p_i\}$. Here, the median filter has a fixed window length $2m+1$, and the values of $\hat{\rho}_{0|1} = \hat{\rho}_{1|0}$ are allowed to vary The left side of the plot represents situations where $e_i = 1$, and the majority of the states in the observation window $\mathbf{e}_i$ are ones. Thus, $m+1 \leq m_1 \leq 2m+1$. The right side of the plot represents situations where $e_i = 0$, and the majority of states in the observation window are zeroes. Thus, $m+1 \leq m_0 \leq 2m+1$. The implicit assumption in the performance plots is that the state run lengths are always greater than $2m+1$. If a state run length were less than $m+1$, the impulse filtering property of the median filter would tend to suppress recognition of such a short state run. As a result, when the state run lengths grow small relative to $m+1$, the error probabilities asymptotically approach unity.

Fig. 4.12 examines the effect of various values $\hat{\rho}_{0|1} = \hat{\rho}_{1|0}$ with the filter length fixed at $2m+1 = 9$. Error probabilities of the smoothed sequence increase monotonically with increasing probability that $\hat{p}_i$ is in error.

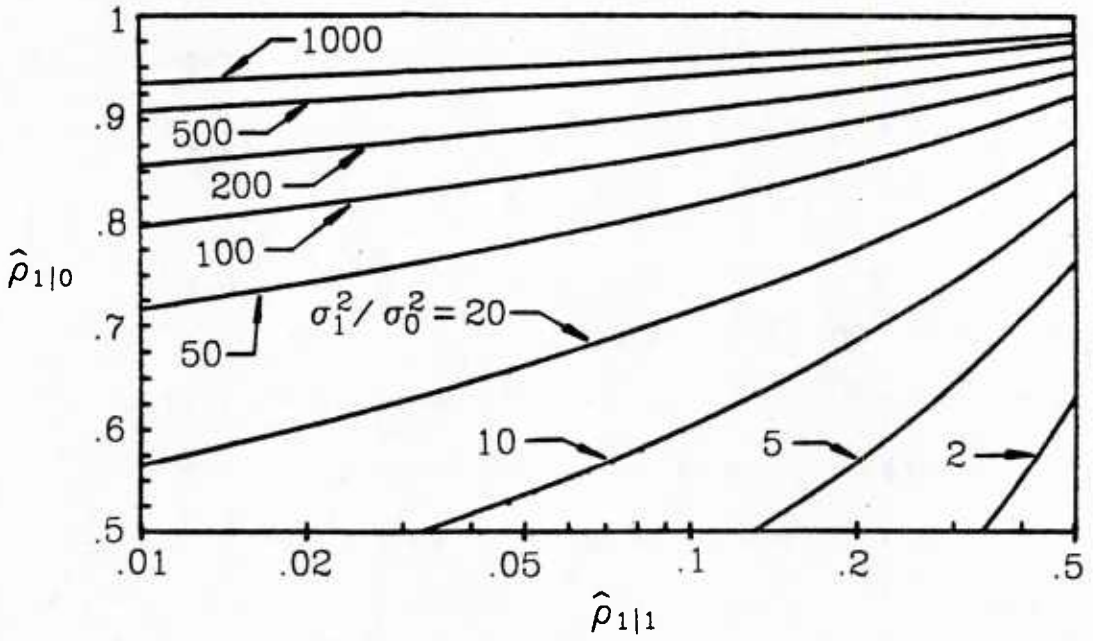

Fig. 4.11. Operating characteristic for the threshold test giving $\hat{p}_i$ for Gaussian-Gaussian switched burst noise. Note that performance is not a function of ε.

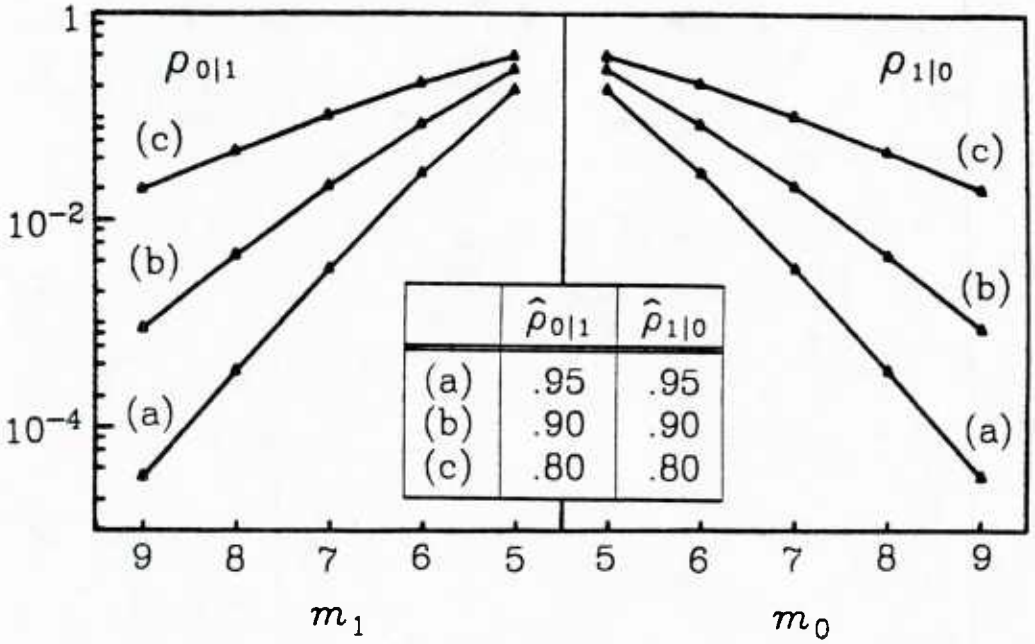

Fig. 4.12. Error performance of smoothed sequence $\{p_i\}$ evaluated for various threshold test error probabilities and $\hat{\rho}_{0|1} = \hat{\rho}_{1|0}$ when $\mathbf{e}_i$ contains a state transition. Filter length is $2m+1 = 9$.

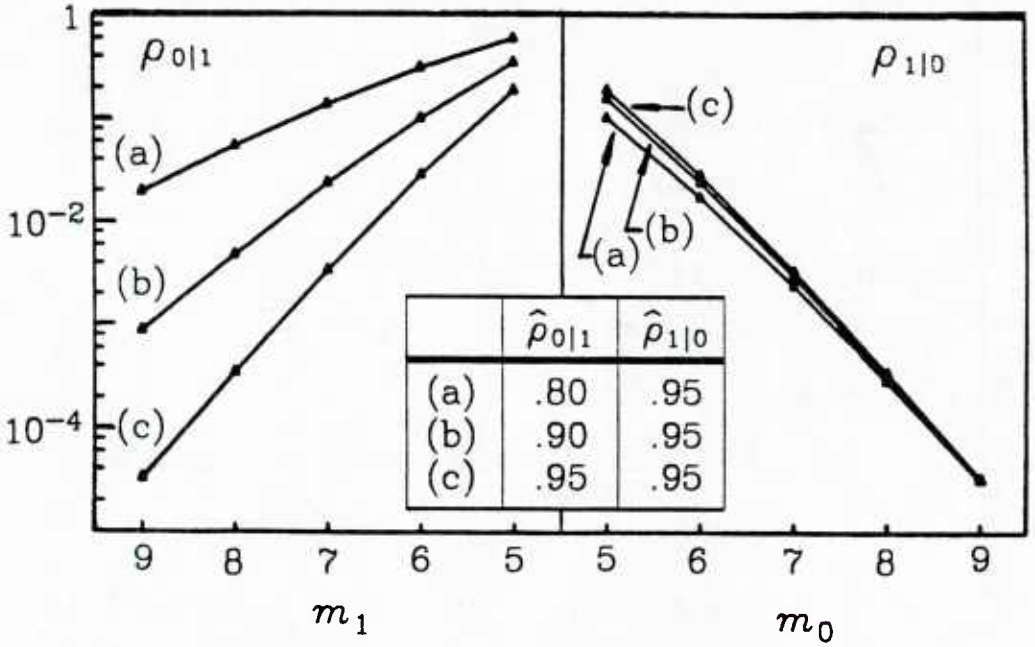

Fig. 4.13. Error performance of smoothed sequence $\{p_i\}$ evaluated for various threshold error probabilities when $\mathbf{e}_i$ contains a state transition. Here, the effect $\hat{\rho}_{0|1} \neq \hat{\rho}_{1|0}$ may be seen. Filter length is $2m+1 = 9$.

Notice that when e_i is near a state transition, $m_0, m_1 \approx m+1$, and correct reconstruction of a state value becomes orders of magnitudes more difficult than when e_i is not near a transition and $m_0, m_1 \approx 2m+1$. When no state transitions occur within the filter window, $\mathbf{e}_i$ is either **0** or **1**. This condition will be denoted as *steady state*, and the probability that p_i is incorrectly classified is at a minimum. The steady state error probabilities are the quantities

$$\rho_{1|0}\Big|_{m_0=2m+1} \tag{4.26}$$

and

$$\rho_{0|1}\Big|_{m_1=2m+1} \tag{4.27}$$

The effect of varying $\hat{\rho}_{0|1}$ while keeping $\hat{\rho}_{1|0}$ fixed is examined in Fig. 4.13. By symmetry, conclusions from this case may be applied to the complementary situation. Unexpectedly, increasing $\hat{\rho}_{0|1}$ decreases $\rho_{1|0}$. This effect is operative only when e_i is near a state transition, and the effect diminishes as e_i moves away from the state transition. For example, assume $e_i = 0$. Then errors in $\hat{p}_i$ after the $0 \rightarrow 1$ or prior to the $1 \rightarrow 0$ transition in $\mathbf{e}_i$ contribute favorably to the filter test statistic when $e_i = 0$. As the state transition propagates through $\mathbf{e}_i$, there are fewer opportunities for $\hat{p}_i$ to *incorrectly* take on state value zero within the filter window. Thus, larger values of $\hat{\rho}_{0|1}$ tend to diminish the probability of failing to recognize that $e_i = 0$.

The effect of changing the filter window length is examined in Fig. 4.14. Here, $\hat{\rho}_{0|1} = \hat{\rho}_{1|0} = .90$, and $2m+1$ is allowed to vary. As the filter

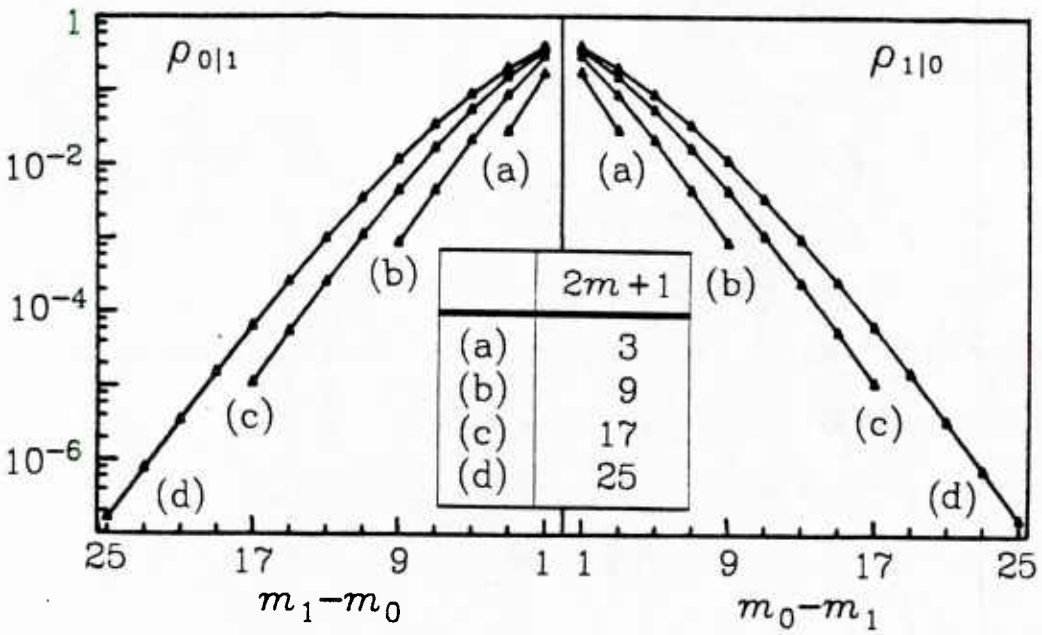

Fig. 4.14. Error performance of smoothed sequence $\{p_i\}$ evaluated for different filter lengths $2m+1$ with $\hat{\rho}_{0|1}=\hat{\rho}_{1|0}=.90$ when $\mathbf{e}_i$ contains a state transition.

length grows, it becomes relatively more difficult near a state transition in $\{e_i\}$ to properly classify p_i. However, this is compensated by the fact that the steady state error probabilities diminish rapidly with increasing filter length. Note that, when $\mathbf{e}_i$ is in steady state, the error probabilities of p_i again are functions only of e_i.

4. Simulation

The algorithm developed in this chapter was applied using the selected high kurtosis Arctic under-ice noise data to simulate a noise source. This data was described in Appendix 2.1 of Chapter 2, and used in the simulations of the previous chapter. As before, the mean of each of the 58 selected blocks was adjusted to zero.

To carry out the test (4.14) and form the sequence $\{\hat{p}_i\}$, the test threshold was chosen as $1.282\hat{\sigma}$, where $\hat{\sigma}^2$ is the variance of each block, calculated as

$$\hat{\sigma}^2 = \frac{1}{1024}\sum_{k=1}^{1024} n_i^2$$

The value $T = 1.282\hat{\sigma}$ corresponds to an error rate of $\hat{p}_{1|0} = .20$ if the noise distribution is indeed Gaussian. If the noise is a background Gaussian noise with a high variance Gaussian impulsive contaminant, then $\hat{\sigma}^2$ overestimates the background variance, and $\hat{p}_{1|0} < .2$. Similarly, $\hat{\sigma}^2$ underestimates the variance of the impulsive component. In typical situations, the impulsive component is present for only a small proportion of the time, and the variance ratio $\sigma_1^2/\sigma_1^2 \gg 1$. Therefore, while a threshold of $1.282\hat{\sigma}$ would correspond to an error rate of $\hat{p}_{0|1} = .80$ if the

noise samples belonged exclusively to the impulsive mode, it is far more likely that $.80 \gg \hat{p}_{0|1}$ under the stated conditions.

With the threshold set at $T = 1.282\hat{\sigma}$ the sequence $\{\hat{p}\}$ was formed. Various window lengths were used to smooth $\{\hat{p}_i\}$, and the best overall estimated value of $ARE_{SB,ld}$ (described later) was obtained with a window length $2m+1 = 7$. Fig. 4.15 presents a representative block of noise data and the corresponding subsequence of $\{p_i\}$.

The non-Gaussian nature of the noise distribution is demonstrated in Fig. 4.16, a Q-Q plot of the empirical distribution of a sample noise data block versus the unit variance Gaussian distribution. In this plot, a Gaussian sample distribution would appear as a straight line. For noise samples near the mean, the plot is approximately linear. For large samples, the empirical noise distribution has a spread greater than that of the Gaussian distribution. Thus, it may be concluded that the noise sample is heavier-tailed than a Gaussian density.

Since the smoothed switching sequence $\{p_i\}$ estimates $\{e_i\}$ and classifies each noise sample as either a background or impulsive noise process observation, the noise samples may be segregated, and the variances $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ may be estimated. Using these estimates and the sequence $\{p_i\}$, each noise sample in the observation block may be normalized to unit variance. Fig. 4.17 presents the data of Fig. 4.15 after this adjustment. Distinct spikes no longer appear in the plot, save for a single spike near sample 800, where $\{p_i\}$ may be in error. The Q-Q plot of the normalized data is shown in Fig. 4.18. The resulting plot is more nearly a straight line, indicating that the normalized data is now

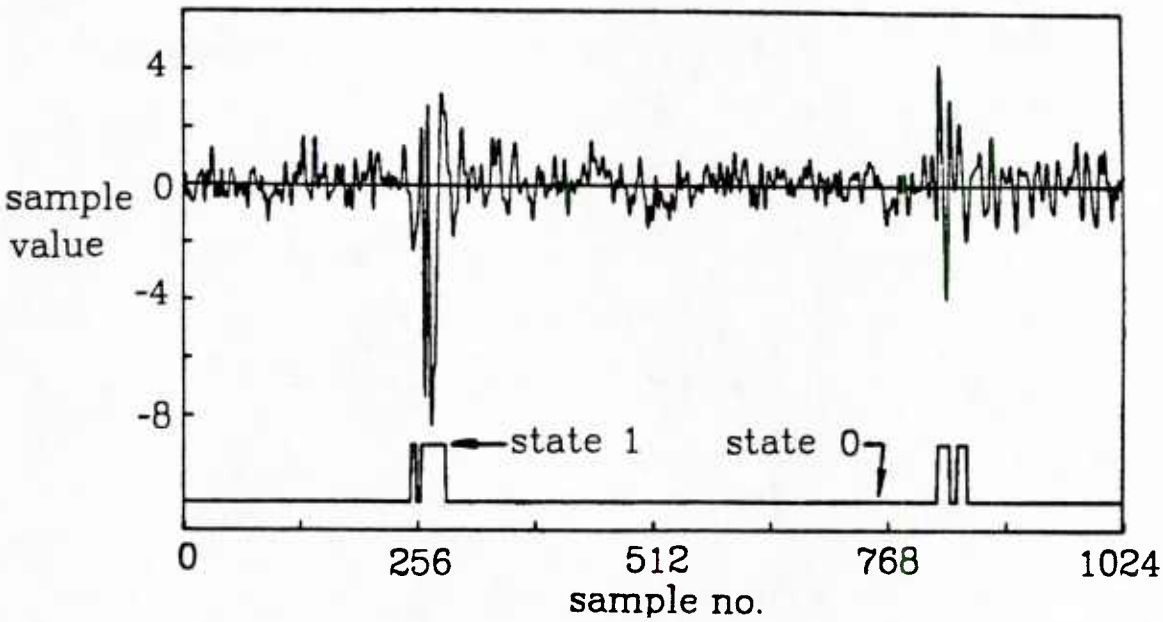

Fig. 4.15. Comparison of sample Arctic under-ice data record 2220 and corresponding subsequence of $\{p_i\}$. Vertical scale of the noise is in standard deviations from the mean. A threshold of $T = 1.282\hat{\sigma}$ and filter length $2m+1 = 7$ were used.

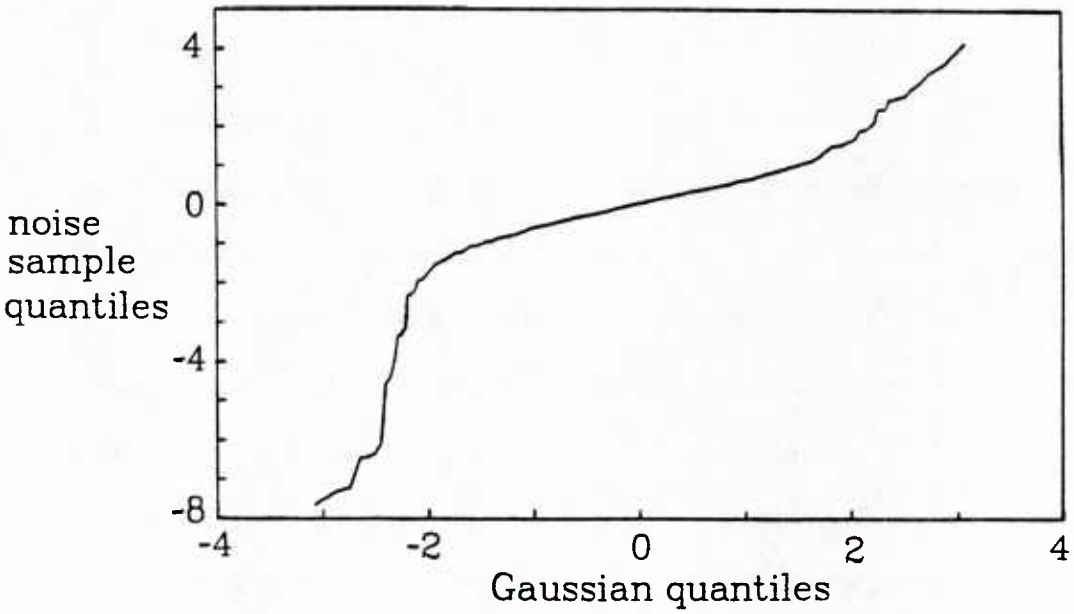

Fig. 4.16. Q-Q plot showing sample quantiles of sample Arctic under-ice data record 2220 prior to processing versus the quantiles of a zero mean unit variance Gaussian distribution.

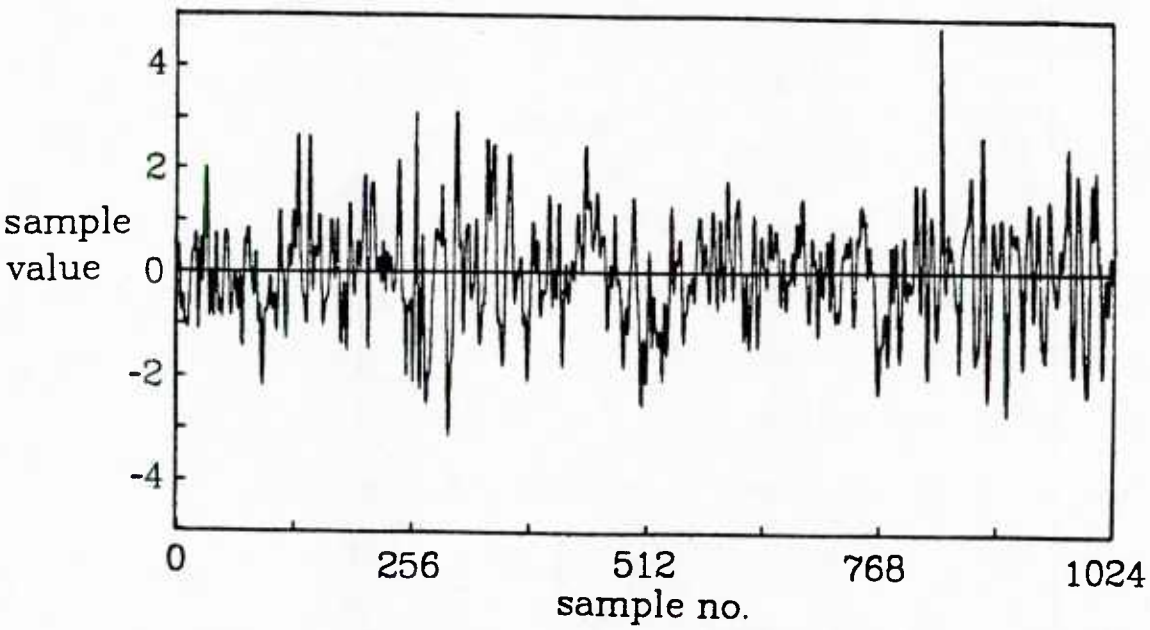

Fig. 4.17. Normalized sample Arctic under-ice noise data record 2220 after $\{p_i\}$ and $\hat{\sigma}_1^2/\hat{\sigma}_0^2$ are used to adjust the data. Vertical scale is in standard deviations from the mean.

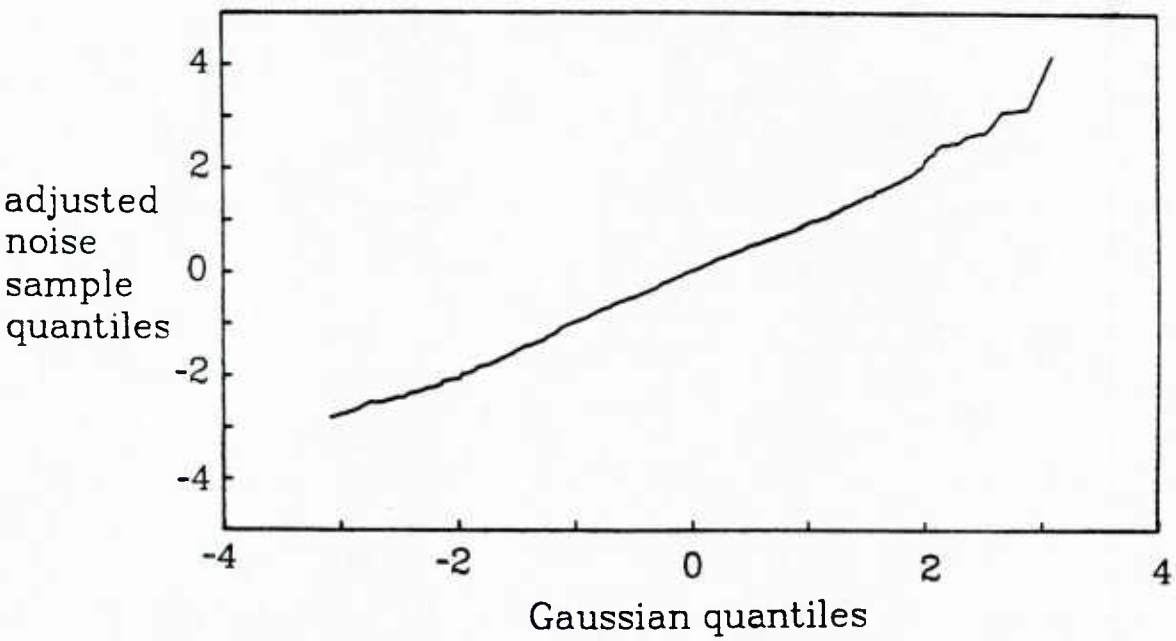

Fig. 4.18. Q-Q plot showing sample quantiles of normalized Arctic under-ice noise data record 2220 after $\{p_i\}$ and $\hat{\sigma}_1^2/\hat{\sigma}_0^2$ are used to adjust the data versus the quantiles of a zero mean unit variance Gaussian distribution.

Gaussian. Therefore, we may conclude that the algorithm provided an effective means of distinguishing between the background and impulsive noise samples.

For the particular data block of Figs 4.15 - 4.17, approximately 5.6% of the samples are classified as impulsive noise, and the variance ratio is estimated as $\hat{\sigma}_1^2/\hat{\sigma}_0^2 = 31.3$. These estimates may be used in (4.9) to estimate η_{SB} by simply setting $\sigma_0^2 = 1$, and letting $\sigma_1^2 = \hat{\sigma}_1^2/\hat{\sigma}_0^2$. In this case, $ARE_{SB,ld} = 2.55$ is the estimated performance improvement.

Using the switched burst detector algorithm, all 58 of the high kurtosis data blocks may be analyzed, allowing ε, and $\hat{\sigma}_1^2/\hat{\sigma}_0^2$ to be estimated for each data block. Fig. 4.19 gives the estimate of ε, and the estimate of the variance ratio is given in Fig. 4.20. Fig. 4.21 presents the values of $ARE_{SB,ld}$ for each data block derived by substituting these estimates into (4.9).

Over the 58 data blocks the cumulative average parameters were computed, giving $\varepsilon = .089$, and $\hat{\sigma}_1^2/\hat{\sigma}_0^2 = 9.03$. These parameter values lead to $ARE_{SB,ld} = 1.58$ as an estimate of the processing gain.

The switched burst detector may be compared to the adaptive detectors of the previous chapter. Fig. 4.22 shows $ARE_{SB,ld}$ plotted with $ARE_{g_{tm},ld}$ and $ARE_{g_{2l},ld}$. Here, it is clear that the switched detector outperforms the non-switching adaptive detectors. This result is not unexpected, for the results of this chapter indicate that g_{SB} outperforms g_ε in Gaussian-Gaussian switched burst noise, and the last chapter indicates that g_ε slightly outperforms both g_{tm} and g_{tm} in Gaussian-Gaussian ε-

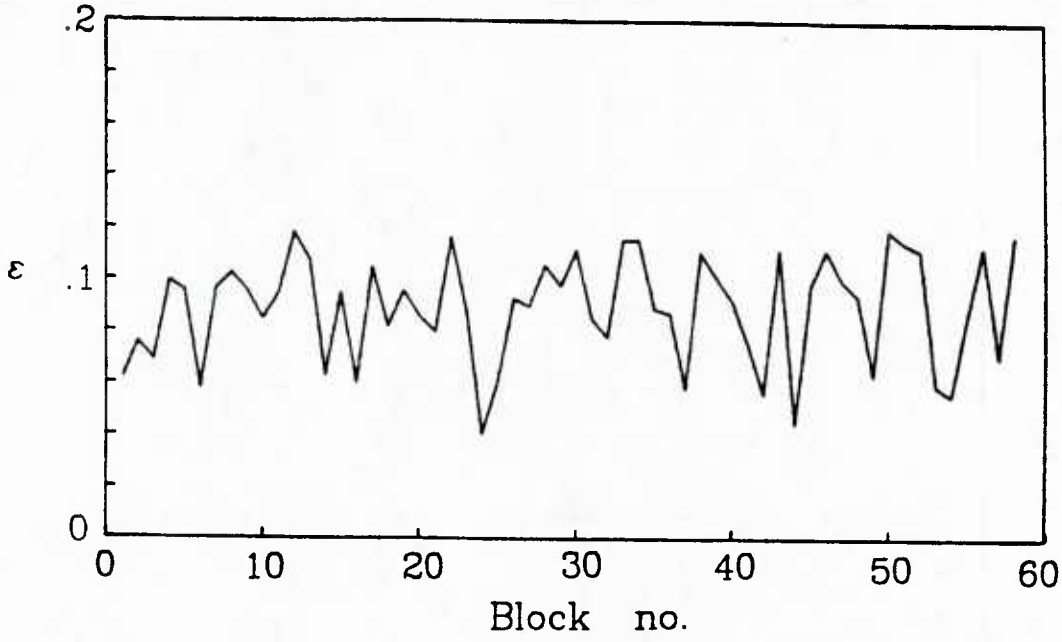

Fig. 4.19. Estimated ε for each sample noise data block.

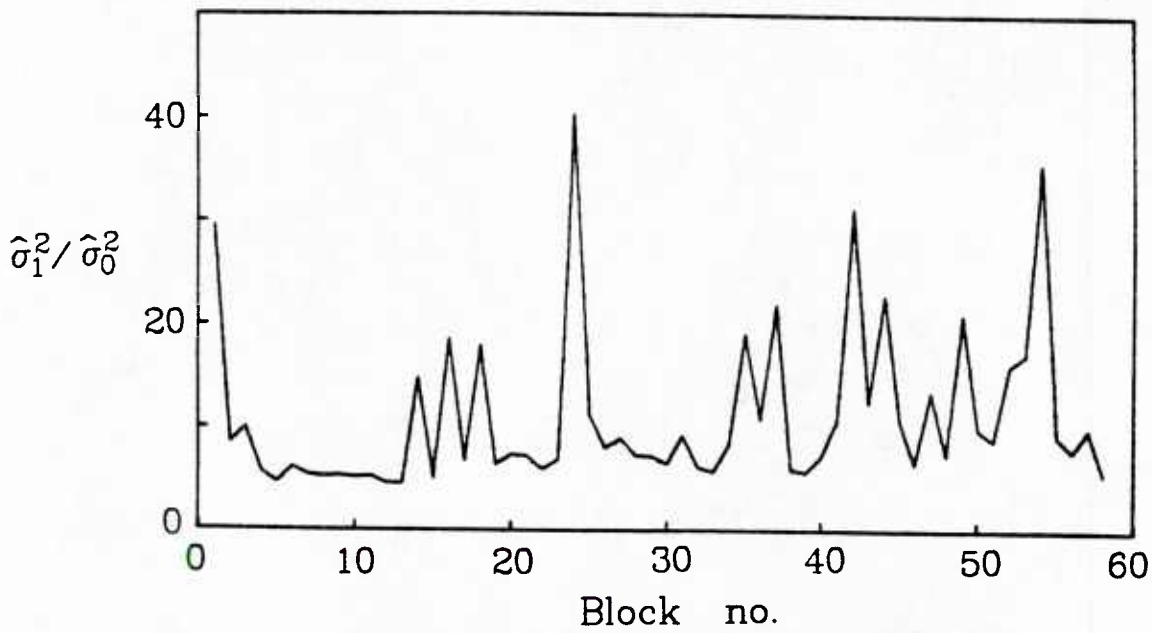

Fig. 4.20. Estimate of $\hat{\sigma}_1^2/\hat{\sigma}_0^2$ for each sample noise data block.

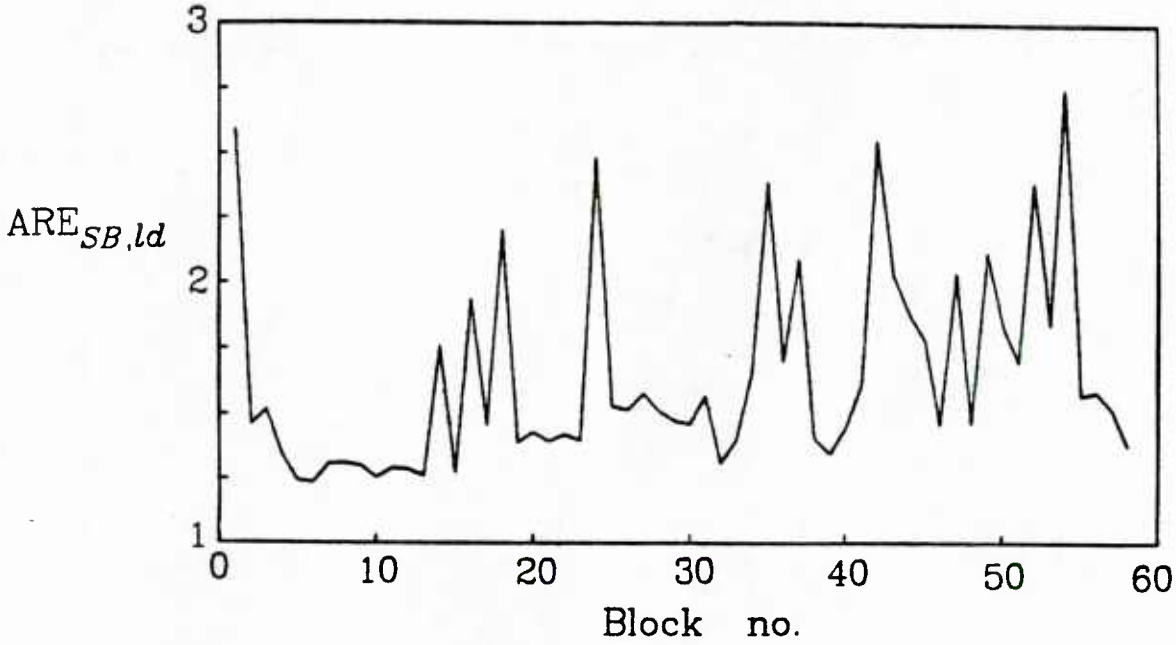

Fig. 4.21. Estimated performance of switched burst detector *SB* relative to a linear detector *ld* for each sample noise data block.

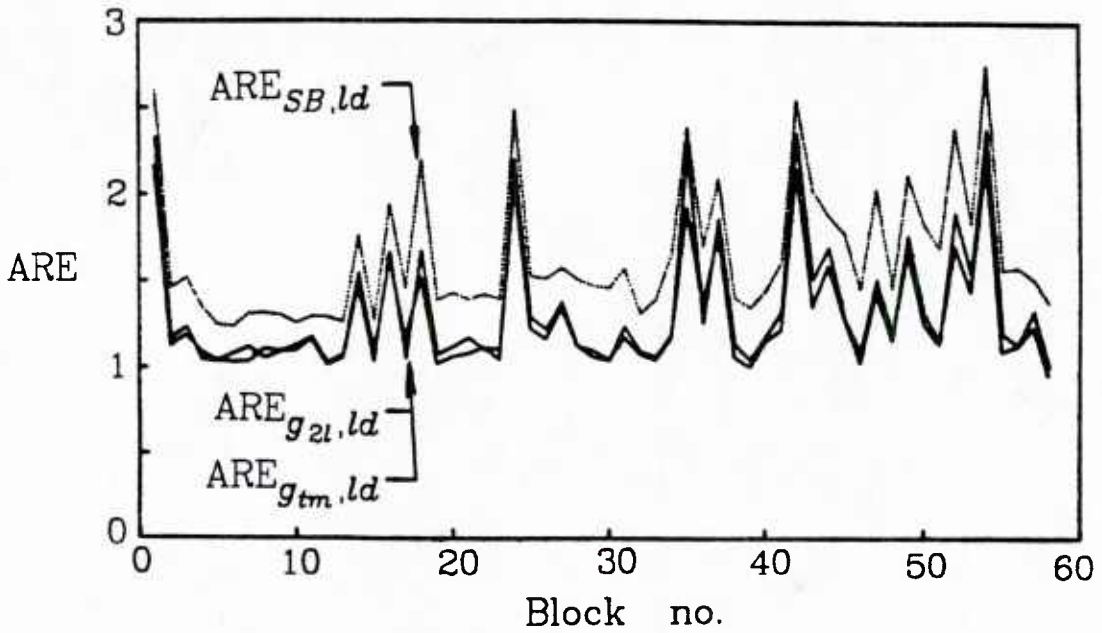

Fig. 4.22. Estimated performance of the switched burst detector *SB* relative to a linear detector *ld* (broken line) with the estimated performance of nonlinearities g_{tm} and g_{2l} relative to *ld* (solid lines) for each sample noise data block.

mixture noise. As noted earlier, the expression for the efficacy of a fixed detector in Gaussian-Gaussian ε-mixture noise is identical to the expression for efficacy in Gaussian-Gaussian switched burst noise when $g_0 = g_1 = g$, where g is some fixed arbitrary detector. It follows then, that g_{SB} will outperform any fixed detector nonlinearity if the Arctic under-ice noise is indeed a Gaussian-Gaussian switched burst noise.

5. Conclusion

We have presented an argument in favor of a time-varying nonlinearity for use in a LO detector structure when the signal is embedded in a type of impulsive noise that classified here as a *switched burst* noise. The nonstationarity part of the structure is easy to implement: it merely requires switching the observations between two fixed nonlinearities. Analysis of the algorithm indicates that this detector is capable of improved performance over a fixed structure. A simple technique for determining the presence of noise bursts has also been proposed, and its performance was examined.

It may be argued that the additional complexity of two nonlinearities and a structure to estimate the switching sequence $\{e_i\}$ is not warranted by the relatively modest improvement over the fixed nonlinearity g_ε. Several points are in order: First, a complex nonlinearity is replaced by two linear amplifiers and a switch. Second, the exact shape of g_ε is a function of the impulsive proportion ε and the variance ratio σ_1^2/σ_0^2. In the proposed algorithm, $\{e_i\}$ would be determined by observation of the noise behavior, and the exact value of ε is of no importance to the

switched detector. The ratio of impulsive to background variance is important, however, since it will determine the gain ratio of the two linear amplifiers. This is easy to calculate, since $\{e_i\}$ separates the noise observations into a stream of observations from the impulsive noise process and a stream of observations from the background noise process. As a result, the two variances may be calcalated in a straigtforward and appealingly natural manner.

An assumption made in the example was that the impulsive component could be adequately modeled by a high variance Gaussian density. It might be desirable to use some other heavy tailed noise to model the impulsive component, for instance, a Laplace density. This has some intuitive appeal: It may be assumed that the impulsive component itself may be modeled with an additive mixture density in a fashion similar to Huber [18]. Then, as the contamination parameter approaches unity, the mixture density approaches the Laplace density, whose LO nonlinearity is the sign detector. Thus, g_0 would be a linear detector, and g_1 a sign detector. Alternatively, g_1 might be chosen to be an amplifier-limiter, a noise blanker, or some other nonlinearity that gives the test statistic a degree of robustness against impulsive noise bursts.

One interpretation of the proposed structure is that switching between two linear detectors is not necessary. Instead, the proposed structure could be regarded as a linear processor with some sort of automatic gain control, which can quickly and accurately adjust an amplifier gain and hold the noise variance constant. A linear detector with continuously adjustable gain is equivalent to the limiting case $M \to \infty$

of a switched burst detector where the detector switches between M linear amplifiers.

Appendix 4.1

In this appendix, an expression is developed for the efficacy of a detector that switches between two nonlinearities g_0 and g_1 in accordance with a control sequence e_i. For the sake of compactness, E_{H_0} denotes the n-fold expectation with respect to the density $\prod_{i=1}^{n} f^i(x_i)$, and E_{f_0} denotes univariate expectation with respect to f_0.

The efficacy of a detector using test statistic T_n is defined by [5,6,9] as

$$\eta_{T_n} = \lim_{n\to\infty} \frac{\left[\frac{\partial}{\partial s} E_{H_1}(T_n)\Big|_{s=0}\right]^2}{n \operatorname{var}_{H_0} T_n} \tag{A4.1}$$

A regularity condition causes the signal s to vanish asymptotically, ensuring that the probability of detection does not converge to unity as n grows without bound [5,9]. Another interpretation is that (A4.1) is an incremental signal-to-noise ratio [6,7], and the regularity conditions guarantee that as $n\to\infty$, the incremental SNR remains finite.

The test statistic for the switched burst detector can be written as

$$T_n = \sum_{i=1}^{n} g^i_{SB}(x_i) \tag{A4.2}$$

$$= \sum_i [(1-e_i)g_0(x_i) + e_i g_1(x_i)]$$

since e_i takes on only the values of zero or unity. Then

$$E_{H_1} T_n = \sum_i \int_{-\infty}^{\infty} [(1-e_i)g_0(x_i+s) + e_i g_1(x_i+s)] \tag{A4.3}$$

$$\times [(1-e_i)f_0(x_i) + e_i f_1(x_i)]dx_i$$

We use the fact that $(1-e_i)^2 = (1-e_i)$ and $e_i^2 = e_i$ and that $e_i(1-e_i) = 0$ to obtain

$$E_{H_1} T_n = \sum_i (1-e_i) E_{f_0} g_0(x_i+s) \qquad \text{(A4.4)}$$

$$+ \sum_i e_i E_{f_1} g_1(x_i+s)$$

Finally, making the usual assumption that the order of expectation and differentiation may be interchanged, we have

$$\frac{\partial}{\partial s} E_{H_1} T_n = E_{f_0} g_0' \sum_{i=1}^{n} (1-e_i) + E_{f_1} g_1' \sum_{i=1}^{n} e_i \qquad \text{(A4.5)}$$

Without loss of generality, we will assume g_0 has zero mean under f_0, and g_1 has zero mean under f_1. Then $\text{var}_{H_0} T_n = E_{H_0} T_n^2$. Here,

$$E_{H_0} T_n^2 = E_{H_0} \left\{ \sum_{i=1}^{n} [(1-e_i)g_0 + e_i g_1] \right\}^2 \qquad \text{(A4.6)}$$

The summands in (A4.6) may be rearranged to obtain

$$E_{H_0} T_n^2 = E_{H_0} \sum_i \left\{ (1-e_i) g_0^2(x_i) + e_i g_1^2(x_i) \right\} \qquad \text{(A4.7)}$$

$$+ E_{H_0} \sum_{i \neq j} \sum \left\{ [(1-e_i)g_0(x_i) + e_i g_1(x_i)] \right.$$

$$\left. \times [(1-e_j)g_0(x_j) + e_j g_1(x_j)] \right\}$$

The sequence $\{x_i\}$ is independent, and g_0 and g_1 are memoryless transformations; therefore the second expectation on the right side of (A4.7) equals zero. Thus,

$$E_{H_0} T_n^2 = E_{f_0} g_0^2 \sum_i (1-e_i) + E_{f_1} g_1^2 \sum_i e_i \tag{A4.8}$$

Substituting (A4.5) and (A4.7) into (A4.1), we find

$$\eta_{T_n} = \lim_{n \to \infty} \frac{\left[\sum_i (1-e_i) E_{f_0} g_0' + \sum_i e_i E_{f_1} g_1'\right]^2}{n\left[\sum_i (1-e_i) E_{f_0} g_0^2 + \sum_i e_i E_{f_1} g_1^2\right]} \tag{A4.9}$$

After multiplying through by n^{-2} and taking the limit, we have

$$\eta_{T_n} = \frac{\left[(1-\varepsilon) E_{f_0} g_0' + \varepsilon E_{f_1} g_1'\right]^2}{(1-\varepsilon) E_{f_0} g_0^2 + \varepsilon E_{f_1} g_1^2} \tag{A4.10}$$

Appendix 4.2

The previous appendix formulated the efficacy for the switched burst detector assuming perfect knowledge of the switching sequence $\{e_i\}$. In this appendix the result (A4.10) is extended to account for errors.

Suppose e_i is the true sequence, but errors are made randomly in choosing between detectors g_0 and g_1. Let $p_i \in \{0,1\}$ represent the decision at observation time i to choose g_0 or g_1, respectively. In the ideal case, $p_i = e_i$ for $-\infty < i < \infty$. To model the effect of errors, let

$$\text{Prob}\,(p_i = 1 \mid e_i = 0) = \rho_{1|0} \tag{A4.11}$$

$$\text{Prob}\,(p_i = 0 \mid e_i = 1) = \rho_{0|1} \tag{A4.12}$$

be the posterior probabilities of determining p_i incorrectly, where the posterior probabilities of correct detection are given by

$$\rho_{0|0} = 1-\rho_{1|0} \tag{A4.13}$$

$$\rho_{1|1} = 1-\rho_{0|1} \tag{A4.14}$$

Clearly, it is desirable to have $\rho_{1|0}$ and $\rho_{0|1}$ as near zero as possible. From the point of view of the detection system, $\{e_i\}$ is a deterministic sequence, and $\{p_i\}$ is a noisy estimate of $\{e_i\}$. It is assumed that (A4.11-A4.14) are time invariant.

Rather than repeat the derivation of Appendix 4.1, only the significant modifications in the derivation will be noted. The correct decision sequence $\{e_i\}$ in (A4.2) and its sequels may be replace by the corrupted decision sequence $\{p_i\}$. Thus, the test statistic T_n becomes

$$T_n = \sum_i \Big[(1-p_i)g_0(x_i)+p_i g_1(x_i)\Big] \tag{A4.15}$$

The expectation of p_i may be taken with respect to its posterior distribution:

$$\text{if } e_i = 1, \quad \text{then } E_{p_i|e_i} = \sum_{k=0}^{1} k \, \text{Prob}(p_i = k \mid e_i = 1) = \rho_{1|1}$$

$$\text{if } e_i = 0, \quad \text{then } E_{p_i|e_i} = \sum_{k=0}^{1} k \, \text{Prob}(p_i = k \mid e_i = 0) = \rho_{0|0}$$

Therefore

$$E_{p_i|e_i} p_i = e_i \rho_{1|1} + (1-e_i)\rho_{1|0} \tag{A4.17}$$

$$E_{p_i|e_i}(1-p_i) = e_i \rho_{0|1} + (1-e_i)\rho_{0|0} \tag{A4.18}$$

Applying these results, the expectation of T_n with respect to the posterior distribution of p_i may be written as

$$E_{p_i|e_i} T_n = \sum_i (1-e_i)\left[\rho_{0|0} g_0(x_i) + \rho_{0|1} g_1(x_i)\right] \tag{A4.16}$$

$$+ \sum_i e_i \left[\rho_{0|1} g_0(x_i) + \rho_{1|1} g_1(x_i)\right]$$

Following the same arguments as in Appendix 4.1,

$$\frac{\partial}{\partial s} E_{H_1} E_{p_i|e_i} T_n = E_{f_0}\left[\rho_{0|0} g_0' + \rho_{0|1} g_1'\right] \sum_{i=1}^{n} (1-e_i) \tag{A4.19}$$

$$+ E_{f_1}\left[\rho_{0|1} g_0' + \rho_{1|1} g_1'\right] \sum_{i=1}^{n} e_i$$

To compute $\text{var}_{H_0} T_n$, the arguments in Appendix 4.1 are paralleled, giving

$$E_{p_i|e_i} T_n^2 = \sum_i E_{p_i|e_i}\left[(1-p_i) g_0^2(x_i) + p_i g_1^2(x_i)\right]$$

$$+\sum_{i\neq j}\sum E_{p_i,p_j|e_i,e_j}\Big[(1-p_i)g_0(x_i)+p_i g_1(x_i)\Big]\Big[(1-p_j)g_0(x_j)+p_j g_1(x_j)\Big]$$

The single summation becomes

$$\sum_i (1-e_i)\Big[\rho_{0|0}g_0^2(x_i)+\rho_{1|0}g_1^2(x_i)right)+e_i\Big[\rho_{0|1}g_0^2(x_i)+\rho_{1|1}g_1^2(x_i)\Big]$$

Depending on the reconstruction algorithm, $\{p_i\}$ may or may not be an independent sequence, so the double summation term cannot be dropped after expectation with respect to $p_i|e_i$. However, every term contains a cross product $g(x_i)g(x_j)$ with $i\neq j$. Thus, the expectation with respect to H_0 of each term in the double summation is zero, for $\{x_i\}$ is an independent sequence, and g_0 and g_1 are memoryless. Therefore,

$$E_{H0}E_{p_i|e_i}T_n^2 = E_{f_0}\Big[\rho_{0|0}g_0^2+\rho_{1|0}g_1^2\Big]\sum_i(1-e_i) \qquad \text{(A4.20)}$$

$$+\,E_{f_1}\Big[\rho_{0|1}g_0^2+\rho_{1|1}g_1^2\Big]\sum_i e_i$$

Following a similar computation as (A4.9), it may be concluded that

$$\eta_{T_n} = \frac{\Big[(1-\varepsilon)E_{f_0}\Big[\rho_{0|0}g_0'+\rho_{1|0}g_1'\Big]+\varepsilon E_{f_1}\Big[\rho_{0|1}g_0'+\rho_{1|1}g_1'\Big]\Big]^2}{(1-\varepsilon)E_{f_0}\Big[\rho_{0|0}g_0^2+\rho_{1|0}g_1^2\Big]+\varepsilon E_{f_1}\Big[\rho_{0|1}g_0^2+\rho_{1|1}g_1^2\Big]} \qquad \text{(A4.21)}$$

Equivalently, noting (A4.13) and (A4.14), the expression for efficacy may be rewritten to depend only on the error probabilities

$$\eta_{T_n} = \frac{\Big[(1-\varepsilon)E_{f_0}\Big[(1-\rho_{1|0})g_0'+\rho_{1|0}g_1'\Big]+\varepsilon E_{f_1}\Big[\rho_{0|1}g_0'+(1-\rho_{1|0})g_1'\Big]\Big]^2}{(1-\varepsilon)E_{f_0}\Big[(1-\rho_{1|0})g_0^2+\rho_{1|0}g_1^2\Big]+\varepsilon E_{f_1}\Big[\rho_{0|1}g_0^2+(1-\rho_{0|1})g_1^2\Big]} \qquad \text{(A4.22)}$$

References

[1] M. Basseville, "Change in Statistical Models: Various Approaches in Automatic Control and Statistics", *Rapport de Recherche, no. 145*, Institut de Recherche en Informatique et Systemes Aleatoires, (In English), March 1981. Also submitted to *IEEE Transactions on Automatic Control.*

[2] A.S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems", *Automatica*, vol. 12, pp. 601-611, 1976.

[3] R.F. Dwyer, "FRAM II Single Channel Ambient Noise Statistics", *Technical Document 6583*, Naval Underwater Systems Center, New London, CT, November,1981. (presented at 101st Meeting, Acoustical Society of America, Ottawa, Canada, May 19, 1981.)

[4] K.S. Vastola, "On Narrowband Impulsive Noise", *Proc. 20th Annual Allerton Conference on Communications, Control, and Computing*, Sept. 1982.

[5] J. Capon, "On the Asymptotic Relative Efficiency of Locally Optimum Detectors", *IRE Trans. on Info. Theory*, vol IT-7, pp. 67-71, April 1961.

[6] C.W. Helstrom, *Statistical Theory of Signal Detection*, Pergamon Press: Oxford, England, 1968.

[7] A. Traganitis, "Narrow Band Filtering, Estimation and Detection", *Ph.D. Dissertation*, Dept. of EECS, Princeton University, Princeton NJ, 1974.

[8] D.F. Andrews, *et al.*, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, NJ, 1972.

[9] E.J.G. Pitman, *Some Basic Theory for Statistical Inference*, Oxford: Pergamon Press, 1968.

[10] D. Middleton, "Statistical-Physical Models of Electromagnetic Interference", *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106-127, August 1977.

[11] D. Middleton, "Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance", *IEEE Trans. Electromagn. Compat.*, vol. EMC-21, no. 3, pp. 209-220, August 1979.

[12] D.R. Halverson and G.L. Wise, "Discrete-time Detection in φ-mixing Noise", *IEEE Trans. on Info. Theory*, vol. IT-26, no. 2, pp. 189-198, March 1980.

[13] H. V. Poor and J.B. Thomas, "Memoryless Discrete-time Detection of a Constant Signal in m-dependent Noise", *IEEE Trans. on Info. Theory*, vol. 25, no. 1, pp. 54-61, Jan. 1979.

[14] P.F. Swaszek, A.B. Martinez, and J.B. Thomas, "Noise Models for Detection", *Proc. IEEE Intl. Conf on Commun.*, Philadelphia, PA, pp. 1H.3.1-1H.3.5, June 13-17, 1982.

[15] S. Cambanis and B. Liu, "On the Expansion of a Bivariate Distribution and its Relationship to the Output of a Nonlinearity", *IEEE Trans. on Info. Theory*, vol IT-17, no. 1, pp. 17-25, Jan. 1971.

[16] J.H. Miller and J.B. Thomas, "Detectors for Discrete-time Signals in Non-Gaussian Noise", *IEEE Trans. on Info. Theory*, vol. IT-18, no. 2, pp. 241-250, March 1972.

[17] J.H. Miller and J.B. Thomas, "The Detection of Signals in Impulsive Noise Modeled as a Mixture Process", *IEEE Trans. on Commun.*, vol. COM-24, no. 5, pp. 559-563, May 1976.

[18] P.J. Huber, "A Robust Version of the Probability Ratio Test", *Ann. Math. Statist.*, vol. 36, pp. 1753-1758, Dec. 1965.

[19] E.N. Gilbert ,"Capacity of a Burst-Noise Channel", *Bell System Tech. Journal*, vol. 39, no. 5, pp. 1253-1265, Sept. 1960.

[20] L. Ehrman, L.B. Bates, J.F. Eschle, and J.M. Kates, "Real-Time Software Simulation of the HF Radio Channel", *IEEE Trans. Commun.*, vol. COM-30, no. 8, pp. 1809-1817, August 1982.

[21] A.M. Mood, F.A. Grayhill, and D.C. Boes, *Introduction to the Theory of Statistics, Third Edition*, McGraw-Hill: New York, 1974.

[22] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. I*, John Wiley and Sons, Inc.: New York, 1968.

[23] N.C. Gallagher and G.L. Wise, "A Theoretical Analysis of the Properties of Median Filters", *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-29, no. 6, pp. 1136-1141, Dec. 1981.

[24] L.A. Berry, "Understanding Middleton's Canonical Formula for Class A Noise", *IEEE Trans. Electromagn. Compat.*, vol. EMC-23, no. 4, pp. 337-343, Nov. 1981.

5

Approximation of Locally Optimum Detector Nonlinearities†

An interesting problem arising in detection is the following: given that the true noise density f and the true detector nonlinearity g_{opt} are known, what is the best way to approximate g_{opt} within some specified constraints? This chapter provides one possible solution to this broadly posed question.

Section 1 reviews the theoretical background of the problem, and states the objective more precisely. Section 2 presents a theorem and proof showing the equivalence of a minimum mean square error (minimum MSE) approximation approach and an efficacy maximizing approach. Section 3 provides some numerical examples as illustration of

† This chapter is based on work done in collaboration with K.S. Vastola of Princeton University; a different version of this chapter appeared as a coauthored paper [15].

the theorem. A summary of the chapter is presented in Section 4.

1. Introduction and Problem Statement

As discussed in Chapter 2, the locally optimal (LO) detector structure is useful for the detection of a signal which is known, but very small relative to the noise environment. For detecting a (constant) weak discrete-time signal in the presence of white non-Gaussian noise with first-order density f, it is well known that the LO detector consists of a memoryless nonlinearity (ZNL) of the form

$$g_{LO}(x) = -\frac{f'(x)}{f(x)} \tag{5.1}$$

followed by summation and comparison with a threshold.

Obviously, when the functional form of f is known explicitly, it is possible to calculate the exact form of g_{LO}. However, it may not be appropriate to implement the exact function g_{LO}; instead, it may be desirable to implement some suboptimal nonlinearity $\hat{g}$. Possible reasons for this may be that $\hat{g}$ is in some sense easier to implement or more easily adaptable to changing noise environments. For instance, $\hat{g}$ may be a ZNL with a simple parameterization. Other considerations may be that the best estimate of g_{LO} (*e.g.*, via density estimates) is too rough or has no closed form representation.

When dealing with weak signal detectors, the usual measure of performance is efficacy [1-4], which can be defined by the following equation

$$\eta_f(g) = \frac{E_f^2(g')}{E_f(g^2)} \tag{5.2}$$

where E_f is the expectation with respect to f. Without loss of generality, we assume $E_f(g) = 0$. The efficacy (5.2) can also be thought of as an incremental signal-to-noise ratio or as the processing gain achievable using detector nonlinearity g when the noise has density f. In principle, the problem discussed above may be solved by maximizing (5.2) over the family of possible ZNL's which we choose to admit. Unfortunately, in practice this is not often a simple thing to do, and an alternative approach is sought.

2. Theorem and Discussion

The theorem presented below yields a method for finding the best nonlinearity over a class of suboptimum nonlinearities. Basically the theorem states that this problem is equivalent to that of finding the nonlinearity which is closest to g_{LO} in the mean square sense. Several related results have been obtained in recent years. For the specific problem of designing detector quantizers, Kassam [5] and Poor and Alexandrou [6] have shown that a close relationship exists between maximum-efficacy quantization and quantization minimizing the mean square distortion relative to g_{LO}. Also, in the more general setting of strong mixing (dependent) noise, Halverson and Wise [7] have shown that if a sequence of nonlinearities $\{\hat{g}_n\}$ converges in mean square to g_{LO}, then the efficacies $\{\eta_f(\hat{g}_n)\}$ converge to the optimal efficacy $\eta_f(g_{LO})$. Note that "mean square", as used in this context, is with respect to the measure defined by the noise distribution.

Within the problem setting of Section 1, the following theorem is a generalization of the results in [5] and [6] discussed above.

Theorem. Given a noise density f, its LO nonlinearity g_{LO}, and a family **G** of candidate suboptimum nonlinearities, the solution $\hat{g}^* \in \mathbf{G}$ to

$$\eta_f(\hat{g}^*) = \max_{\hat{g} \in \mathbf{G}} \eta_f(\hat{g}) \tag{5.3}$$

is the same as the solution $\hat{g}^* \in \mathbf{G}$ to

$$E_f(\hat{g}^* - g_{LO})^2 = \min_{\hat{g} \in \mathbf{G}} E_f(\hat{g} - g_{LO})^2. \tag{5.4}$$

subject to a simple normalization of the elements in **G**.

Proof. Under the mild conditions of the Pitman-Noether Theorem [1,3]

$$\eta_f(g) = \frac{\left[\int g(x) f'(x) dx\right]^2}{\int g^2(x) f(x) dx} \tag{5.5}$$

Our problem is: Given a class **G** of nonlinearities and a density f, find $g^* \in \mathbf{G}$ solving

$$\max_{g \in \mathbf{G}} \eta_f(g) \tag{5.6}$$

Since the efficacy of a nonlinearity g is invariant under a scale change (*i.e.*, $\eta_f(cg) = \eta_f(g)$ for every $c \neq 0$), we can multiply each nonlinearity g by the constant

$$c_g = -\left[\frac{\int g_{LO}^2(x) f(x) dx}{\int g^2(x) f(x) dx}\right]^{1/2} \operatorname{sgn}\left[\int g(x) f'(x) dx\right] \tag{5.7}$$

This allows us to assume, without loss of generality, that for every $g \in \mathbf{G}$

$$\int g^2(x)f(x)dx = \int g_{LO}^2(x)f(x)dx \tag{5.8}$$

and

$$\int g(x)f'(x)dx \leq 0 \tag{5.9}$$

Now consider the MSE problem

$$\min_{g \in \mathbf{G}} \int (g(x)-g_{LO}(x))^2 f(x)dx \tag{5.10}$$

We have straightforwardly that

$$\int (g(x)-g_{LO}(x))^2 f(x)dx$$

$$= \int g^2(x)f(x)dx \tag{5.11}$$

$$-2\int g(x)g_{LO}(x)f(x)dx + \int g_{LO}^2(x)f(x)dx$$

From (5.8) and (5.1) the MSE becomes

$$= 2\left[\int g_{LO}^2(x)f(x)dx + \int g(x)f'(x)dx\right]$$

Because of (5.9), we see that minimizing this over **G** is equivalent to maximizing $\left[\int g(x)f'(x)dx\right]^2$. By (5.8) the quantity $\int g^2(x)f(x)dx$ is constant over **G**; thus we have the conclusion that minimizing the MSE (5.10) is equivalent to maximizing the efficacy functional given in (5.5). ■

Discussion

Thus, given f and g_{LO}, as well as **G**, a family of approximations, the nonlinearity $\hat{g}^*$ which maximizes efficacy over the family **G** is simply the minimum mean square error approximation to g_{LO} over **G**. Solving the

minimum MSE problem (5.4) is often easier than solving (5.3) directly, especially when **G** is a parameterized family.

For the purposes of the proof, each element in **G** was multiplied by the constant c_g, but in practice it is not always necessary to precondition each member of **G**. If one were trying to solve the MSE problem over a parameterized family of nonlinearities, say, $\mathbf{G}=\{\hat{g}(x;\mathbf{a})\}$, with **a** an m-vector of parameters, the simplest approach is to merely treat c_g as an additional parameter controlling the scaling of $\hat{g}$. The new problem then would be to find the minimum MSE estimate of g_{LO} in $\mathbf{G}=\{c_g\hat{g}(x;\mathbf{a})\}$ where the new parameterization is the $(m+1)$-vector $(c_g,\mathbf{a})$. If an explicit amplitude parameter is already an element of **a**, then this modification is unnecessary and (5.4) may be solved directly.

The theorem provides support for certain intuitive ideas about suboptimal detection. Previous work with suboptimal structures [8-14] suggests that near optimal efficacy is possible if the suboptimal structure $\hat{g}$ appears "close to" g_{LO}. Further refinements making $\hat{g}$ "closer" to g_{LO} yield only minor improvements in performance. Since efficacy is directly related to the mean square error between $\hat{g}$ and g_{LO}, it is easy to see that small errors in $\hat{g}$ (relative to g_{LO}) tend to be deemphasized, at the expense of emphasizing the gross errors. Furthermore, the square errors are weighted by the noise density; for unimodal densities, points in the tail region are weighted much less heavily than those near the mode.

These points illustrate why a great deal of latitude is available to the designer in choosing the tail behavior of $\hat{g}$, while the shape of the central

region must be chosen much more carefully. In particular, for heavy-tailed noises, reasonable performance levels may be attained by carefully matching the shapes of $\hat{g}$ and g_{LO} near the noise mean, and choosing more roughly the limiting or blanking behavior of the tail regions [8-14]. Also, note that the adaptive nonlinearities of Chapter 2 typically were good matches to g_{LO} near the noise mean, but only loosely approximated the tails of g_{LO}. In the examples given, these suboptimal adaptive nonlinearities achieved high levels of performance with respect to the optimal nonlinearity. Additionally, in Chapter 3, the only nonlinearities that were substantially suboptimal were cases in which there was a poor fit near the origin.

3. Examples

Known Density

Since maximizing efficacy is the same as solving the MSE problem, the best approximation in **G** is the projection of g_{LO} onto **G**. As an illustration of this point, suppose **G** is the span of a finite set of basis functions φ_i, with $i = 1, \ldots, N$, where the φ_i are orthonormal with respect to f. An approximation $\hat{g}$ will take the form

$$\hat{g} = \sum_{i=1}^{N} a_i \varphi_i \tag{5.12}$$

where the a_i are not all zero. Solving (5.3) directly requires simultaneous solution for $\{a_i\}$ in an N-dimensional quadratic form. Solving (5.4) leads to the solution $a_i = \mathrm{E}_f(g_{LO}\varphi_i)$ for $i = 1, \ldots, N$.

This approach is probably most useful in an analytical context, for

detailed knowledge of f is necessary to generate the orthonormal basis set. If f is not available, a set of N basis functions may still be generated provided $2N$ moments of f are known [17].

Unknown Density

In this example, the theorem is applied to smooth an estimate of g_{LO}. Using a finite number of noise observations, $\{X_i\}_{i=1}^{N}$, the kernel density estimation procedure of Parzen and Rosenblatt [16,18,19] is used to give $\hat{f}$ and $\hat{f}'$, estimates of the density and its first derivative. The LO nonlinearity may then be estimated as $\hat{g}_{LO}(x) = -\hat{f}'(x)/f(x)$. Unless N is very large, $\hat{g}_{LO}$ will be rough, and it will be desirable to find $\tilde{g}_{LO}$, a smoothed version of the estimated nonlinearity. By the theorem, a smoothing technique based on a minimum MSE criterion would yield the best performing $\tilde{g}_{LO}$.

Consider the following numerical example, where the $\{X_i\}$ are 100 iid observations of a zero mean, unit variance noise process with Gaussian-Gaussian ε-mixture density, $\varepsilon = 0.1$, and $\sigma_1^2/\sigma_0^2 = 100$. Using the finite width polynomial kernel and window sizing procedure discussed by Silverman [16], both $\hat{f}$ and $\hat{f}'$ were estimated, and $\hat{g}_{LO}$ was computed. Figures 5.1 and 5.2 compare $\hat{f}$ to the true density.

To smooth $\hat{g}_{LO}$, it was projected onto $\mathbf{G} = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \times (1, x, x^2, x^3)$. In a practical problem, f is unknown, so the expectations are computed with respect to the empirical cdf. Solving the MSE problem (5.4) requires the simultaneous solution of four linear equations. The result is a smoothed estimate

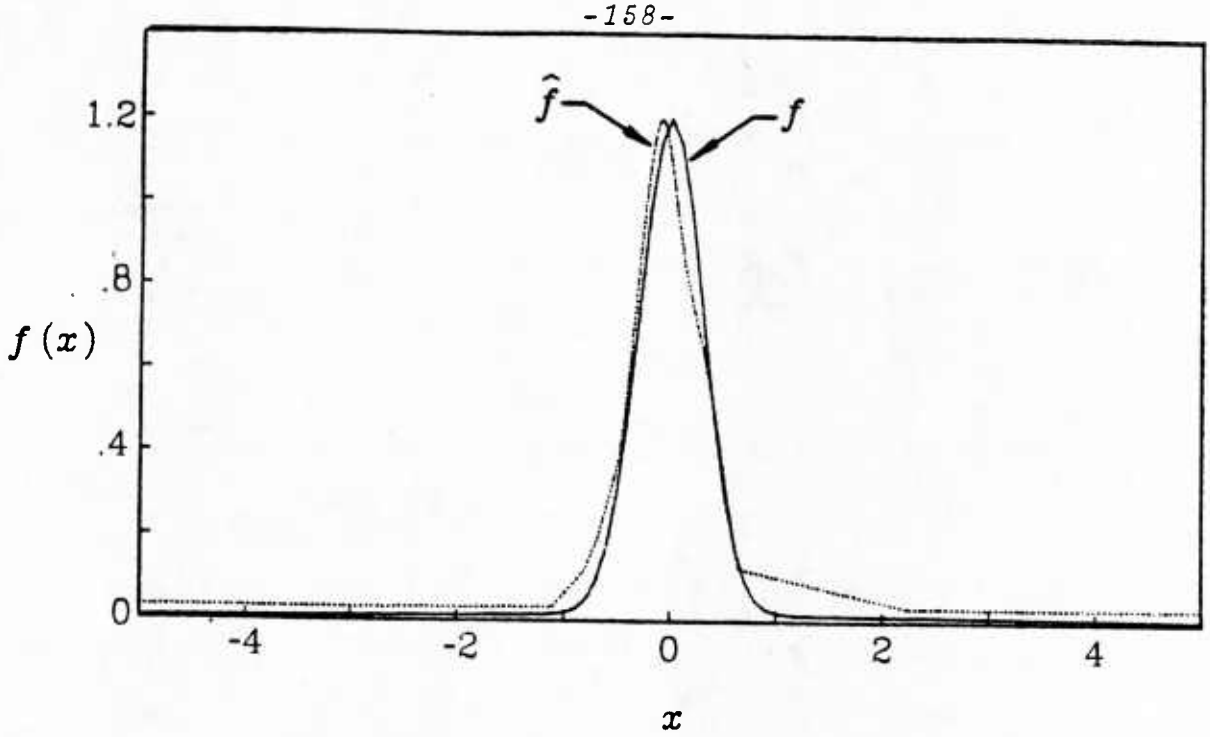

Fig 5.1. Estimated density $\hat{f}$ (broken line) and the true Gaussian-Gaussian ε-mixture density f (solid line).

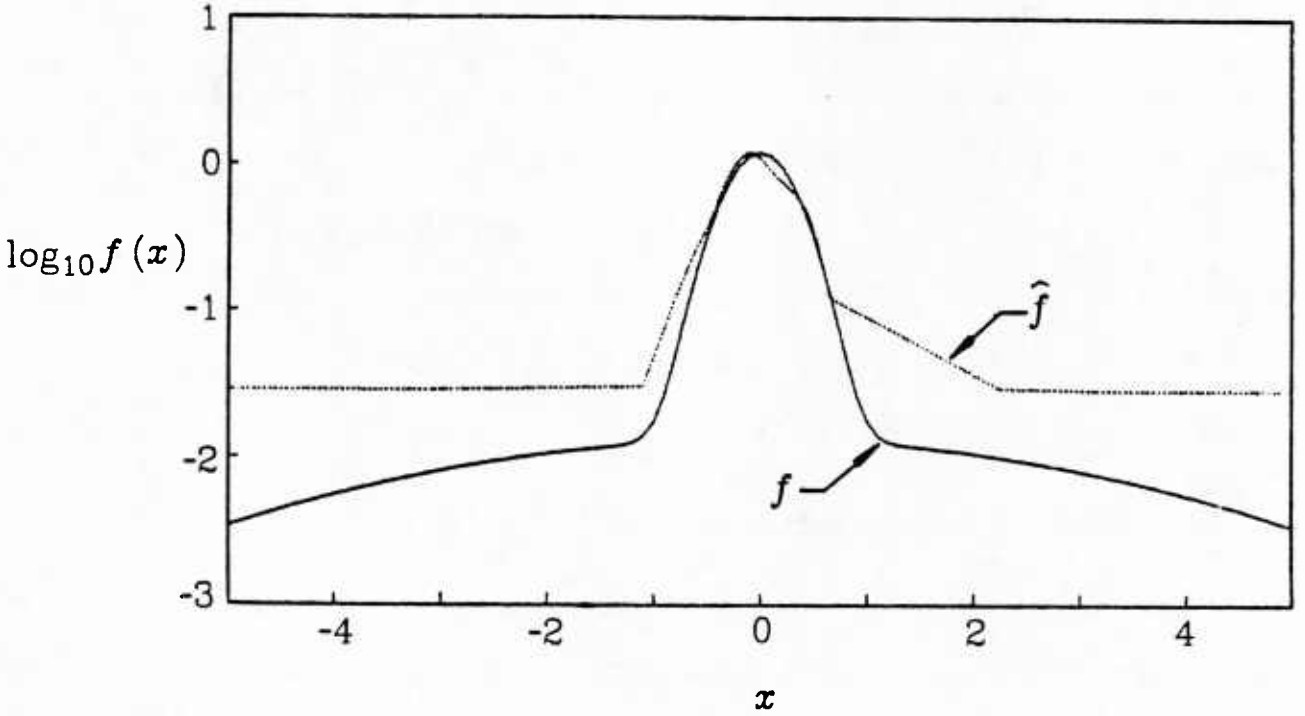

Fig. 5.2. Comparison of $\hat{f}$ (broken line) and f (solid line) on a logarithmic scale.

$$\tilde{g}_{LO}(x) = (\beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta) \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

Figure 5.3 compares $\hat{g}_{LO}$ and $\tilde{g}_{LO}$, and Figure 5.4 compares $\tilde{g}_{LO}$ and the true LO nonlinearity g_{LO}. For this example, $ARE_{\tilde{g}_{LO},ld} = 8.58$ and $ARE_{\tilde{g}_{LO},g_{LO}} = .951$, where ARE is as defined in Chapter 2.

In this example, **G** is not orthonormal with respect to the noise density. It was chosen for convenience and "nice" smoothness properties. This example, and work by Modestino [20], suggest that elements of **G** could be various generic detector nonlinearities, where the coefficients β_i would weight the contribution of each nonlinearity. Some adaptive procedure could observe the noise and update the coefficients β_i.

4. Conclusion

When replacing a known locally optimal nonlinearity with some suboptimal nonlinearity, it is desirable to have a method which is simple and generates a nonlinearity which preserves a high performance level. We have presented a proof of the equivalence of efficacy maximization and mean square error approximation. MSE minimizing procedures have many appealing properties, and they have a rich history in both theory and application. Often relatively simple algorithms may be found for carrying out the calculations, and it is possible that these methods may now be applied fruitfully to the problem of designing maximum efficacy suboptimal detector nonlinearities.

There are several other useful interpretations of the theorem. The first is that, since the MSE performance measure involves only a single

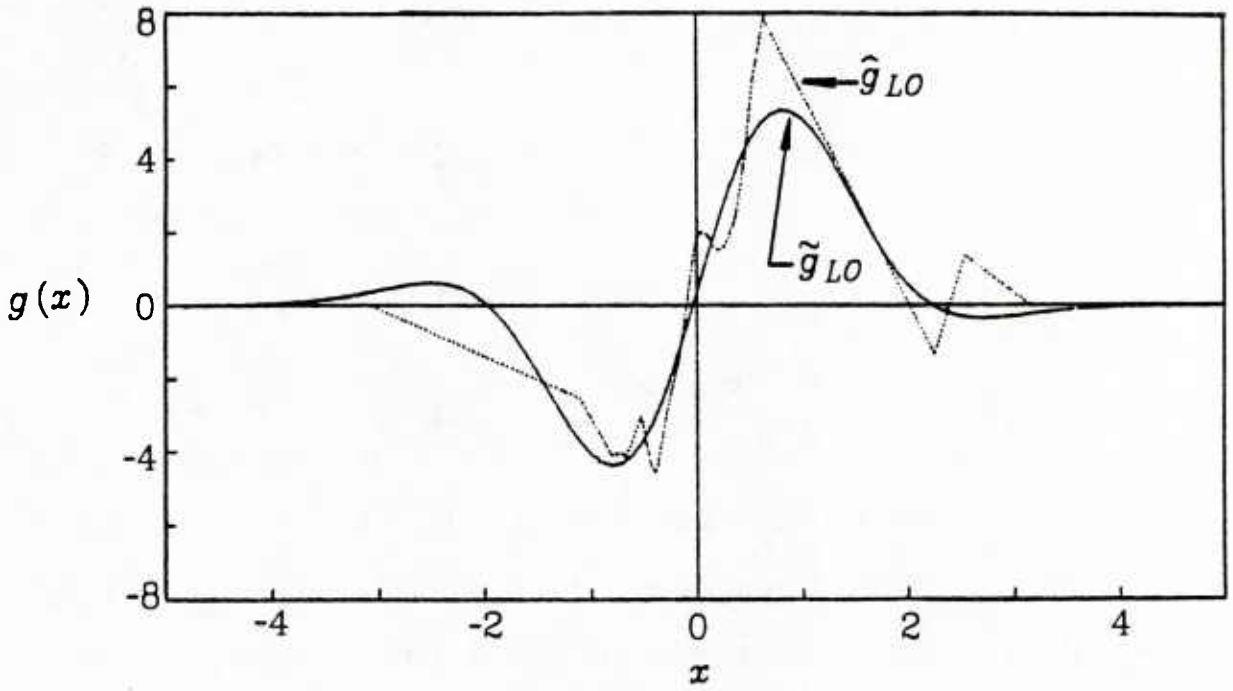

Fig. 5.3. Comparison of the estimated nonlinearity $\hat{g}_{LO}$ (broken line) and the smoothed estimate $\tilde{g}_{LO}$ (solid line).

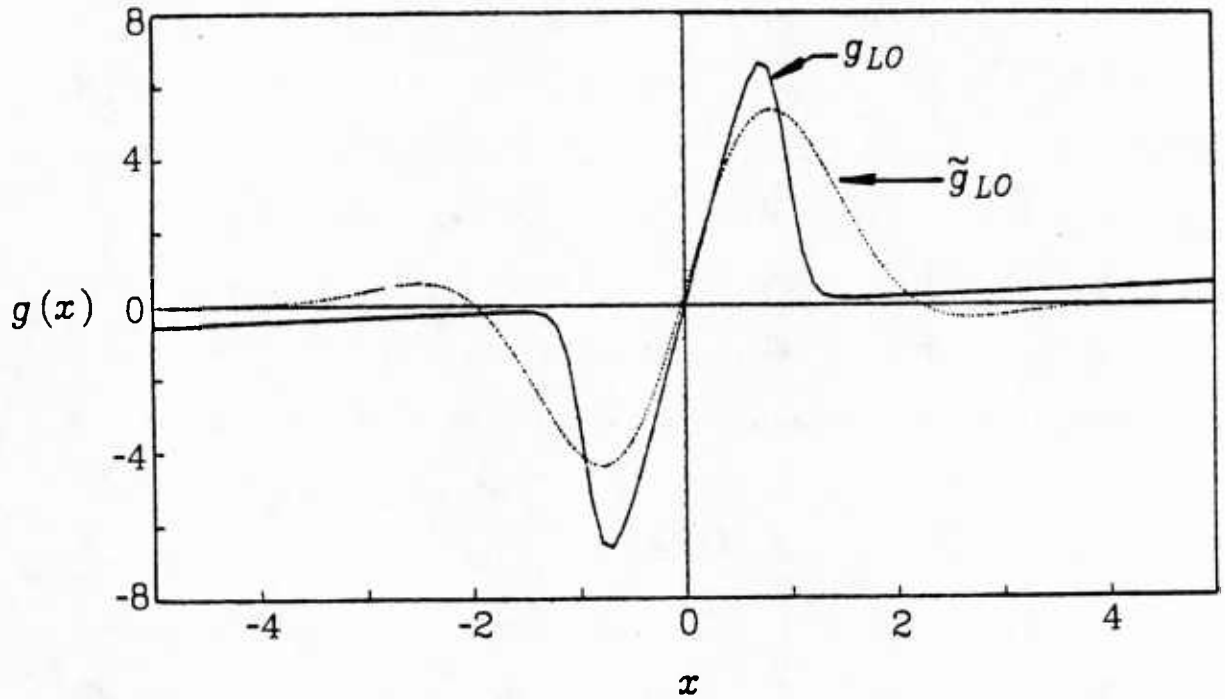

Fig. 5.4. Comparison of the smoothed nonlinearity $\tilde{g}_{LO}$ (broken line) and the true nonlinearity g_{LO} (solid line).

integral, we can study the contribution of an isolated region of the nonlinearity to overall mean square error, and therefore, its relative contribution to performance degradation. As an example, this allows us to examine the sensitivity of performance with respect to changes in the nonlinearity over certain regions of the input axis. Often, the behavior of a nonlinearity's tail region is of particular interest, and the simple ranking of performance sensitivities afforded by the use of (5.10) would allow the relative merits of various tail configurations to be studied independently of the shape of the rest of the nonlinearity. Zero mean square error in the tail region would indicate that the tail is "locally optimum in that region", and therefore provides the best possible contribution to overall performance.

One area of interest still open is the question of approximating the small sample (Neyman-Pearson) detector. It would be worthwhile investigating the properties of minimum MSE approximations to the NP detector nonlinearity.

References

[1] J. Capon, "On the Asymptotic Efficiency of Locally Optimum Detectors," *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67-71, April 1961.

[2] D. Middleton, "Canonically Optimum Threshold Detection," *IEEE Trans. Inform. Theory*, vol. IT-12, No. 2, pp. 230-243.

[3] E.J.G. Pitman, *Some Basic Theory for Statistical Inference*. Chapman and Hall: London, England, 1979.

[4] C.W. Helstrom, *Statistical Theory of Signal Detection*. Pergamon Press: Oxford, England, 1968.

[5] S.A. Kassam, "Optimum Quantization for Signal Detection," *IEEE Trans. Comm.*, vol. COM-25, pp. 479-484, May 1977.

[6] H.V. Poor and D. Alexandrou, "A General Relationship between two Quantizer Design Criteria," *IEEE Trans. Inform. Th.*, vol. IT-26, No. 2, pp. 210-212, March 1980.

[7] D.R. Halverson and G.L. Wise, "On the performance of approximately optimal memoryless detectors for signals in dependent noise," *Proc. 1981 Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD, pp. 147-152, March 1981.

[8] J.H. Miller and J.B. Thomas, "Robust Detectors for Signals in Non-Gaussian Noise," *IEEE Trans. Comm.*, vol. COM-25, pp. 686-690, July 1977.

[9] R.F. Ingram and R.Houle, "Performance of the Optimum and several Suboptimum Receivers for Threshold Detection of Known Signals in Additive, White, Non-Gaussian Noise," Naval Underwater Syst. Dept. Tech. Rept. No. 6339, Newport, Rhode Island.

[10] E.M. Modugno, III, and J.B. Thomas, "Measures of Tail Behavior and their Relationship to Signal Detection," *Proc. 1981 Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD, pp. 26-31, March 1981.

[11] S.V. Czarnecki and J.B. Thomas, "Adaptive Detection of a Known Signal in Non-Gaussian Noise," *Proc. Twentieth Ann. Allerton Conf. Comm., Control, Comput.*, Monticello, IL., pp. 759-768, Oct. 1982.

[12] S.V. Czarnecki and J.B. Thomas, "An Adaptive Detector of non-Gaussian noise", *Proc. IEEE Internat. Conf. Comm.*, Boston, MA, June 1983.

[13] K.S. Vastola, "On Narrowband Impulsive Noise", *Proc. Twentieth Annual Allerton Conf. Comm., Control, Computing*, Monticello, IL, pp. 739-748, Oct. 1982.

[14] K.S. Vastola and S.C. Schwartz, "Suboptimum Threshold Detection in Narrowband Non-Gaussian Noise", *Proc. IEEE Internat. Conf. Comm.*, Boston, MA, June 1983.

[15] S.V. Czarnecki and K.S. Vastola, "Approximation of Locally Optimum Detector Nonlinearities", *Proc. 1983 Conf. Inform. Sci. and Syst.*, Johns Hopkins University, Baltimore, MD, pp. 218-221, March 1983.

[16] B.W. Silverman, "Choosing the Window Width when Estimating a Density", *Biometrika*, vol. 65, no. 1, pp. 1-11, Jan.. 1978.

[17] M.H. Meyers, "Computing the Distribution of a Random Variable via Gaussian Quadrature Rules", *Bell System Tech. Journal*, vol. 61, no. 9, pp. 2245-2261, Nov. 1982.

[18] E. Parzen, "On Estimation of a Probability Density Function and Mode", *Ann. Math. Stat.*, vol. 33, pp. 1065-76, 1962.

[19] M. Rosenblatt, "Remarks on some Nonparametric Estimates of a Density Function", *Ann. Math. Stat.*, vol. 27, pp. 832-837, 1956.

[20] J. W. Modestino, "Adaptive Detection of Signals in Impulsive Noise Environments," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 1022-1026, Sept. 1977.

6

Detection and Small Sample Performance Measurement

Previous chapters have been concerned mainly with locally optimum (LO) detection. As pointed out, LO detection may be regarded as a limiting worst case, optimal only in an asymptotic sense. For finite sample sizes and nonzero signal-to-noise ratios, Neyman-Pearson detection is optimal in a particular sense. Efficacy is a useful asymptotic performance measure, but it does not give much information about the small sample size performance of a detector.

This chapter will be concerned with developing a performance measure useful for comparing finite sample detectors which approximate the NP optimal detector. Section 1 reviews the theoretical background of this problem, and develops the properties of the proposed performance measure, and Section 2 presents some examples applying the result.

Section 3 provides a brief conclusion to the chapter.

1. Analysis of the Performance Index Properties

Introduction and Theoretical Preliminaries

Consider the binary hypothesis testing problem:

$$\begin{aligned} H_0&: \mathbf{x} \sim f_0(\mathbf{x}) \\ & \qquad\qquad \mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{X}^n \\ H_1&: \mathbf{x} \sim f_1(\mathbf{x}) \end{aligned} \tag{6.1}$$

A straightforward application of the Neyman-Pearson Lemma [1, p. 193] leads to a threshold test of the form

$$\Lambda_{NP}(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{6.2}$$

This test is optimal in the sense that for any probability of *false alarm* $\alpha \leq \alpha_0$ of incorrectly deciding H_1 when H_0 is true, the probability β of correctly deciding H_1 when H_1 is true is greater than any other test with level $\alpha \leq \alpha_0$. Often, β is called the *power* of the test. Alternatively, the measure $1-\beta$ is sometimes of interest, and is designated as the probability of *false dismissal*.

As noted in Chapter 2, the statistics α_0 and β are difficult to compute. However, one approach to describing the performance of the test (6.2) is to find bounds on α_0 and $1-\beta$ based upon measures of distance between f_0 and f_1, such as the Chernoff distance [2]. Kailath [3] provides a summary of classical approaches, and Blahut [4] explores distance measures and some connections between hypothesis testing and coding.

An extensive example of Chernoff bounding is available in Van Trees [7, pp 116-133].

These techniques are useful when f_0 and f_1 are known exactly, but unfortunately, this is not often the case. Furthermore, by force or choice, the likelihood ratio test (6.2) may be altered by replacing f_0 and f_1 with incorrect densities p_0 and p_1. Kazakos [5,6] considers the use of distance-measure-like bounding techniques for hypothesis tests based on inaccurate versions of the true densities.

The contribution of this chapter is to extend some results on distance bounding and bounding for detection under mismatch to the more general situation where the likelihood ratio is replaced by a general transformation not necessarily defined by the ratio of two unique densities. It will be useful to make the transformations

$$\lambda_{NP}(\mathbf{x}) = \ln \Lambda_{NP}(\mathbf{x}) \tag{6.3}$$

$$t = \ln T \tag{6.4}$$

and consider the Neyman-Pearson test

$$\lambda_{NP}(\mathbf{x}) \underset{H_0}{\overset{H_1}{\gtrless}} t \tag{6.5}$$

The log-likelihood ratio $\lambda_{NP}(\mathbf{x})$ will be replaced by a general detection processor $g(\mathbf{x})$. The following regularity conditions are assumed with respect to both the measures induced by density functions f_0 and f_1:

(a) $\quad -\infty < g(\mathbf{x}) < \infty$ a.e. in $\mathbf{X}^n$

(b) $f_0(\mathbf{x}) \neq f_1(\mathbf{x})$ for some subset of $\mathbf{X}^n$ with nonzero measure. Thus, distinctness of the hypotheses is assured. Additionally, it is required that the measures induced by f_0 and f_1 both be absolutely continuous with respect to each other. This implies that f_0 and f_1 have common support, and that the detection problem is not singular.

(c) $-\infty < E_0 g < E_1 g < \infty$. Therefore, distinctness of the hypotheses after processing by the detector is assured. This mild condition merely restricts the processor g to be "reasonable": observations under H_1 tend to generate a larger valued test statistic than observations under H_0.

The regularity conditions ensure that $g(\mathbf{x})$ exists w.p.1 under either H_0 or H_1. *It will be assumed that these regularity conditions are satisfied by all detectors and densities considered in the remainder of this chapter.* Using the generalized detection processor g, the likelihood ratio test (6.2) becomes

$$g(\mathbf{x}) \underset{H_0}{\overset{H_1}{\gtrless}} t \tag{6.6}$$

As an aside, note that $g(\mathbf{x})$ has several common realizations. For instance, it may be the output of a matched filter, or its approximation. In other cases, $g(\mathbf{x}) = \sum_{i=1}^{n} g_i(x_i)$, where g_i is a memoryless nonlinear transformation. When the observations $\{x_i\}_{i=1}^n$ are independent

$$\Lambda_{NP}(\mathbf{x}) = \prod_{i=1}^{n} \Lambda_{NP;i}(x_i) \tag{6.7}$$

where $\Lambda_{NP;i}(x_i) = \dfrac{f_{1;i}(x_i)}{f_{0;i}(x_i)}$ is the likelihood ratio of the univariate densities of observation x_i. It follows from the monotonicity of the logarithm function that (6.6) is an NP optimal test when $g_i(x_i) = \lambda_{NP;i}(x_i)$, the log-likelihood ratio. Memoryless transformations other than the log-likelihood ratio may be used as g_i, and they may be generated by methods similar to those proposed in the previous chapters, particularly when the noise density is assumed to be stationary.

Exposition of the Performance Index

For the remainder of this chapter, we consider the the binary hypothesis test of (6.1), assume a decision will be made according to a test (6.6), where the regularity conditions are satisfied and $g(\mathbf{x})$ is not necessarily equal to $\lambda_{NP}(\mathbf{x})$. As a first step in developing a performance index for the test (6.6), consider the functionals given by

Definition 1.

$$M_0(u;g) = \ln \int_{-\infty}^{+\infty} e^{ug(\mathbf{x})} f_0(\mathbf{x}) d\mathbf{x} \tag{6.8}$$

$$M_1(u;g) = \ln \int_{-\infty}^{+\infty} e^{-ug(\mathbf{x})} f_1(\mathbf{x}) d\mathbf{x} \tag{6.9}$$

■

Notice that both M_0 and M_1 are cumulant generating functions, since they are the natural logarithms of the moment generating function (mgf) for the random variables produced by the transformation $g(\mathbf{x})$ or $-g(\mathbf{x})$, respectively. Thus, necessary and sufficient conditions for M to exist and

be finite is that, for u in some neighborhood about the origin, the mgf of g exists. A necessary condition for finite M to exist is simply that $g(\mathbf{x})$ has finite moments of all order.

The following theorem provides bounds on the error probabilities.

Theorem 1. Let a test of the form (6.6) be used to distinguish between two hypotheses of the form (6.1), and assume M_0 and M_1 are defined as above. If M_0 and M_1 exist and are finite, then

$$\alpha_0 \leq e^{(-ut+M_0)} \tag{6.10}$$

$$1-\beta \leq e^{(ut+M_1)} \tag{6.11}$$

Proof. (after [5])

$$\alpha_0 = \text{Prob}\left[G(\mathbf{x}) > T \mid H_0\right]$$

$$= \text{Prob}\left[e^{g(\mathbf{x})} > e^{t} \mid H_0\right]$$

$$= \text{Prob}\left[e^{g(\mathbf{x})-t} > 1 \mid H_0\right]$$

$$\leq \text{E}_0\left[e^{-ut} e^{ug(\mathbf{x})}\right]$$

by the Markov inequality. The proof for the inequality on $1-\beta$ follows in similar manner. ■

The two functionals may be combined to provide a useful performance measure for comparing false dismissal error probabilities of two competing detector structures operating with equal false alarm rates. Before this is illustrated, it is first necessary to develop

Lemma 1. If M_0 and M_1 exist, define the function $r(u;g) = M_0(u;g)+M_1(u;g)$. Then

(*i*)

$$r^*(g) = r(u^*;g) = \left\{M_0(u^*;g)+M_1(u^*;g)\right\} \qquad (6.12)$$

$$= \min_u \left\{M_0(u;g)+M_1(u;g)\right\}$$

exists for some finite u^*.

(*ii*) The minimum value satisfies $r(u^*;g) < 0$.

(*iii*) For any g, the value of u^* is unique.

Proof. †

(*i*) Since M_0 and M_1 are cumulant generating functions, they are convex in u, [8, p. 121]; therefore r is also convex. Observe that $r(0;g) = 0$; it may be shown that $\lim_{u\to\pm\infty} r(u;g) = \infty$.

First, we rewrite the definition of $r(u;g)$ as

$$r(u;g) = \ln e^{-uC}\int e^{u(g+C)} f_0 + \ln e^{uC}\int e^{-u(g+C)} f_1$$

where C is some constant. The region of integration may be partitioned, giving

$$r(u;g) = -uC+uC+\ln\left[\int_{(g+C)>0} e^{u(g+C)} f_0 + \int_{(g+C)\le 0} e^{u(g+C)} f_0\right]$$

† In this proof, we employ *a.e.* as the abbreviation of *almost everywhere*. $f \ll g$ means that the measure induced by f is *absolutely continuous* with respect to the measure induced by g. If $f \ll g$ and $g \ll f$, then the induced measures are *equivalent*, and the condition is denoted as $f \equiv g$. For convenience, the phrase *with respect to the measure induced by* will be suppressed in the text.

$$+ \ln\left[\int_{(g+C)\geq 0} e^{-u(g+C)} f_1 + \int_{(g+C)<0} e^{-u(g+C)} f_1\right]$$

We now consider separately the case $u > 0$. It follows that

$$r(u;g) \geq \ln\left[\int_{(g+C)>0} e^{u(g+C)} f_0\right] + \ln\left[\int_{(g+C)<0} e^{-u(g+C)} f_1\right]$$

as each partitioned integral is nonnegative. Regularity condition (c) implies that g cannot be a constant a.e. with respect to f_0 or f_1. Therefore, for some C, the regularity condition that $f_0 \equiv f_1$ ensures that there exists $\varepsilon > 0$ such that

$$0 \neq \int_{g+C>\varepsilon} f_0$$

and

$$0 \neq \int_{g+C<-\varepsilon} f_1$$

Because $\varepsilon > 0$ exists,

$$r(u;g) \geq \ln\left[\int_{g+C>\varepsilon} e^{u\varepsilon} f_0\right] + \ln\left[\int_{g+C<-\varepsilon} e^{u\varepsilon} f_1\right]$$

$$\geq 2u\varepsilon + \ln\left[\int_{(g+C)>\varepsilon} f_0\right] + \ln\left[\int_{(g+C)<-\varepsilon} f_1\right]$$

The latter function grows without bound as u approaches infinity.

For the case $u < 0$, similar arguments show that $r(u;g)$ grows without bound as u approaches negative infinity also; therefore, since $r(u;g)$ is convex, some finite u^* exists that

minimizes $r(u;g)$.

(*ii*) Since r is convex in u, to show $u^* > 0$ and $r(u^*;g) < 0$, it will be sufficient to demonstrate that $\left.\frac{\partial r}{\partial u}\right|_{u=0} < 0$. Here,

$$\frac{\partial r}{\partial u} = \frac{\partial}{\partial u} M_0(u;g) + \frac{\partial}{\partial u} M_1(u;g)$$

$$= \frac{m_0'(u)}{m_0(u)} + \frac{m_1'(u)}{m_1(u)}$$

where the $m(u)$ are moment generating functions. Thus, $m(0) = 1$, and $m'(0) = \mathrm{E}(g)$, which gives

$$\left.\frac{\partial r}{\partial u}\right|_{u=0} = \mathrm{E}_0(g) + \mathrm{E}_1(-g)$$

Regularity condition (c) ensures that this quantity is negative. Therefore, the minimum value of $r(u;g)$ exists for some $u^* > 0$, and this minimum value $r^*(g)$ is less than zero.

(*iii*) To demonstrate that u^* is unique, it will be sufficient to show that the second partial derivative of $r(u;g)$ with respect to u is strictly positive for all u and arbitrary g. The second partial derivative may be written as

$$\frac{\partial^2 r}{\partial u^2} = \frac{\int e^{ug} f_0 \int g^2 e^{ug} f_0 - \left[\int g e^{ug} f_0\right]^2}{\left[\int e^{ug} f_0\right]^2}$$

$$+ \frac{\int e^{-ug} f_1 \int g^2 e^{-ug} f_1 - \left[\int g e^{-ug} f_1\right]^2}{\left[\int e^{-ug} f_1\right]^2}$$

Notice that the functions

$$f_{0\varepsilon} = \frac{e^{ug} f_0}{\int e^{ug} f_0}$$

and

$$f_{1\varepsilon} = \frac{e^{-ug} f_1}{\int e^{-ug} f_1}$$

are density functions also. The expectations with respect to these two new densities will be denoted by $E_{0\varepsilon}$ and $E_{1\varepsilon}$, respectively. The second derivative may be expressed in terms of these expectations as

$$\frac{\partial^2 r}{\partial u^2} = \left[E_{0\varepsilon} g^2 - E_{0\varepsilon}^2 g\right] + \left[E_{1\varepsilon} g^2 - E_{1\varepsilon}^2 g\right]$$

which is the sum of the variance of g under $f_{0\varepsilon}$ and $f_{1\varepsilon}$, respectively. The regularity conditions ensure that g is finite a.e. and not a.e. a constant; therefore $f_{0\varepsilon} \equiv f_0$, and $f_{1\varepsilon} \equiv f_1$. Then g is not a.e. a constant with respect to $f_{0\varepsilon}$ and $f_{1\varepsilon}$, and its respective variances are nonzero. Thus, when $r(u;g)$ exists, $\frac{\partial^2 r}{\partial u^2}$ is strictly positive. ■

The previous lemma demonstrates that for a given g, it is possible to find the minimum value $r^*(g)$, which shall be designated as a *performance index* of g. The reason for this will be clear from

Theorem 2. If M_0 and M_1 exist, then $1-\beta \le \frac{1}{\alpha_0} e^{(M_0+M_1)}$ for $u > 0$, and there exists a tightest bound

$$1-\beta \le \frac{1}{\alpha_0} e^{r^*(g)} \qquad (6.13)$$

Proof.

$$\alpha_0 \leq \exp(-ut + M_0)$$

$$\ln \alpha_0 \leq -ut + M_0$$

$$t \leq \frac{M_0 - \ln \alpha_0}{u}$$

We substitute this result into the bound on false dismissal probability:

$$1-\beta \leq \exp(ut + M_1)$$

$$\leq \exp(M_0 - \ln \alpha_0 + M_1)$$

$$= \frac{1}{\alpha_0} e^{M_0 + M_1}$$

Lemma 1 guarantees that $r^*(g)$ exists for a unique value $u^* > 0$. Thus, it follows that a tightest bound $\frac{1}{\alpha_0} e^{r^*(g)}$ exists.

■

If $r^*(g)$ is to be a useful performance index for comparing detectors, it must give the best index for the optimal detector structure. Demonstration of this fact will require

Lemma 2. Suppose **G** is a convex set of functions on $\mathbf{X}^n$ satisfying regularity conditions (a) and (c). Suppose M_0 and M_1 exist for all $g \in \mathbf{G}$. Then M_0 and M_1 are convex on **G**.

Proof. To demonstrate convexity of the two functionals it will be sufficient to show that

$$M_0(u; \delta g + [1-\delta]h) \leq \delta M_0(u; g) + [1-\delta] M_0(u; h)$$

for $0 \le \delta \le 1$ and $g, h \in \mathbf{G}$.

We begin by recalling Hölder's Inequality:

$$\text{if } \frac{1}{p}+\frac{1}{q}=1, \text{ then } \mathrm{E}|XY| \le \left(\mathrm{E}|X^p|\right)^{1/p}\left(\mathrm{E}|Y^q|\right)^{1/q}$$

The inequality is applied to the definition of M_0, with $p = \frac{1}{\delta}$ and $q = \frac{1}{1-\delta}$.

$$M_0(u;\delta g+[1-\delta]h) = \ln \int e^{\delta u g(\mathbf{x})} e^{(1-\delta)uh(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x}$$

$$\le \ln \left(\left[\int e^{\delta u g(\mathbf{x})\frac{1}{\delta}} f_0(\mathbf{x})d\mathbf{x} \right]^{\delta} \left[\int e^{(1-\delta)uh(\mathbf{x})\frac{1}{1-\delta}} f_0(\mathbf{x})d\mathbf{x} \right]^{1-\delta} \right)$$

$$= \delta \ln \int e^{ug(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x} + (1-\delta)\ln \int e^{uh(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x}$$

$$= \delta M_0(u;g) + [1-\delta]M_0(u;h)$$

The proof for M_1 is identical in form. ■

Theorem 3. Let $\mathbf{G}$ be the set of all functions on $\mathbf{X}^n$ satisfying (a) and (c). Then the function $r(u;g)$ achieves a globally minimum value for $g(\mathbf{x}) = \lambda_{NP}(\mathbf{x})$ and $u = \frac{1}{2}$.

Proof. First, note that if $g(\mathbf{x}) \in \mathbf{G}$, then $Cg(\mathbf{x}) \in \mathbf{G}$ for any constant $C > 0$. Therefore, minimizing $r(u;g)$ over $\mathbf{R}_+ \times \mathbf{G}$ is equivalent to minimizing $r(\frac{1}{2};g)$ over $\mathbf{G}$. To prove the theorem, it will be sufficient to fix $u = \frac{1}{2}$ and attend to the minimization problem in $\mathbf{G}$.

To prove existence of a stationary point at $g(\mathbf{x}) = \lambda_{NP}(\mathbf{x})$, a calculus of variations argument will be used.

Let $\delta(\mathbf{x})$ be any arbitrary variation which is not a.e. a constant, and let ε be a real number; further, let $\delta(\mathbf{x})$ be subject to the restriction that the perturbed nonlinearity

$$g(\mathbf{x}) = \lambda_{NP}(\mathbf{x}) + \varepsilon\delta(\mathbf{x})$$

remains an element of **G**. If $\delta(\mathbf{x})$ is a.e. a constant, then $r(\tfrac{1}{2};\lambda_{NP}) = r(\tfrac{1}{2};\lambda_{NP}+\varepsilon\delta)$.

The functional $r(u;\lambda_{NP}+\varepsilon\delta)$ may be written

$$r(u;g) = \ln \int e^{u[\lambda_{NP}(\mathbf{x})+\varepsilon\delta(\mathbf{x})]} f_0(\mathbf{x})\,dx$$

$$+ \ln \int e^{-u[\lambda_{NP}(\mathbf{x})+\varepsilon\delta(\mathbf{x})]} f_1(\mathbf{x})\,dx$$

For the remainder of the proof, the dependence on $\mathbf{x}$ will be suppressed in the notation. Taking the first derivative with respect to epsilon yields

$$\frac{\partial r}{\partial \varepsilon} = \frac{u\int \delta e^{u[\lambda_{NP}+\varepsilon\delta]} f_0}{\int e^{u[\lambda_{NP}+\varepsilon\delta]} f_0} - \frac{u\int \delta e^{-u[\lambda_{NP}+\varepsilon\delta]} f_1}{\int e^{-u[\lambda_{NP}+\varepsilon\delta]} f_1}$$

A necessary condition for a stationary point in r to exist at λ_{NP} is that $\left.\frac{\partial r}{\partial \varepsilon}\right|_{\varepsilon=0} = 0$ for all possible variations δ. Therefore, the condition for a minimum is

$$0 = \frac{\int \delta e^{u\lambda_{NP}} f_0}{\int e^{u\lambda_{NP}} f_0} - \frac{\int \delta e^{-u\lambda_{NP}} f_1}{\int e^{-u\lambda_{NP}} f_1}$$

But $\lambda_{NP} = \ln \frac{f_1}{f_0}$, and $u = \tfrac{1}{2}$, and the necessary condition becomes

$$0 = \frac{\int \delta (f_0 f_1)^{1/2}}{\int (f_0 f_1)^{1/2}} - \frac{\int \delta (f_1 f_0)^{1/2}}{\int (f_1 f_0)^{1/2}}$$

which is fulfilled for any arbitrary variation δ. Therefore, the conclusion is that

$$\frac{\partial}{\partial \varepsilon} r(\tfrac{1}{2}; \lambda_{NP} + \varepsilon \delta) \Big|_{\varepsilon=0} = 0$$

It is easy to show that **G** is convex; hence, Lemma 2 implies that this stationary point is a global minimum [13, p. 191].

As an aside, note that the global minimum value is achieved for any pair of u and g such that $ug = \frac{1}{2}\lambda_{NP} + C$ almost everywhere for any constant C. Thus, the globally minimum value $r(\frac{1}{2};\lambda_{NP})$ is not unique. ■

In a binary hypothesis test, the performance of the test is unaffected by a monotone transformation of the test statistic. Here, the weaker property of the invariance of $r^*(g)$ to linear transformations of the test is demonstrated.

Proposition 1. $r^*(g) = r^*(a + bg)$, where the variables a and b are real numbers, and $b \neq 0$.

Proof.

$$r(u;g) = M_0(u;a+bg) + M_1(u;a+bg) \qquad (6.14)$$

$$= \ln \int e^{u(a+bg(\mathbf{x}))} f_0(\mathbf{x}) d\mathbf{x} + \ln \int e^{-u(a+bg(\mathbf{x}))} f_1(\mathbf{x}) d\mathbf{x}$$

$$= \ln e^{ua} \int e^{ubg(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x} + \ln e^{-ua} \int e^{-ubg(\mathbf{x})} f_1(\mathbf{x})d\mathbf{x}$$

$$= ua - ua + \ln \int e^{wg(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x} + \ln \int e^{-wg(\mathbf{x})} f_1(\mathbf{x})d\mathbf{x}$$

$$= M_0(w;g) + M_1(w;g) \qquad (6.15)$$

Finally, minimization of (6.14) with respect to u obviously yields the same result as minimization of (6.15) with respect to w. Therefore $r^*(g)$ is invariant under linear transformations on g. ■

2. Application of the Performance Index

The previous section proposed the performance index $r^*(g)$ and developed some of its properties under very loose regularity conditions on the two hypothetical densities f_0 and f_1, as well as on the detection processor g. The index is usable for dependent as well as independent noise, and for linear or nonlinear processors, with or without memory.

The iid Noise Case

The properties of the index will be explored here for the case of independent and identically distributed observations where g is the summation of outputs of a memoryless nonlinear transformation.

Proposition 2. Let the noise densities of hypothesis test (6.1) be $f_{(\cdot)}(\mathbf{x}) = \prod_{i=1}^{n} f_{(\cdot);i}(x_i)$, and let the detection processor $g(\mathbf{x})$ be of the form $g(\mathbf{x}) = \sum_{i=1}^{n} g_i(x_i)$. Then

$$M_0(u;g) = \sum_{i=1}^{n} \ln \int e^{ug_i(x_i)} f_{0;i}(x_i)dx_i \qquad (6.16)$$

$$M_1(u;g) = \sum_{i=1}^{n} \ln \int e^{-ug_i(x_i)} f_{1;i}(x_i)dx_i \qquad (6.17)$$

Proof. The proof is a straightforward computation, outlined here for M_0 as

$$M_0(u;g) = \ln \int e^{ug(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x}$$

$$= \ln \underbrace{\int \cdots \int}_{n\text{-fold}} \exp\left[u \sum_{i=1}^{n} g_i(x_i)\right] \prod_{i=1}^{n} f_{0;i}(x_i)dx_1 \cdots dx_n$$

$$= \ln \prod_{i=1}^{n} \int e^{ug_i(x_i)} f_{0;i}(x_i)dx_i$$

$$= \sum_{i=1}^{n} \ln \int e^{ug_i(x_i)} f_{0;i}(x_i)dx_i$$

The proof for M_1 follows similarly. ■

When the noise observations are iid, then $f(\mathbf{x}) = \prod_{i=1}^{n} f(x_i)$. Here, the distinction between the multivariate and univariate densities should be clear from the arguments of the densities.

Corollary 1. When the noise is independent and identically distributed, and $g(\mathbf{x}) = \sum_{i=1}^{n} g_1(x_i)$, then

$$M_0 = nM_0(u;g_1) = n \ln \int e^{ug_1(x)} f_0(x)dx$$

$$M_1 = nM_1(u;g_1) = n \ln \int e^{-ug_1(x)} f_1(x)dx$$

■

The performance index becomes $r^*(g) = nr^*(g_1)$ where

$$r^*(g_1) = \min_u \left[M_0(u;g_1) + M_1(u;g_1) \right] \qquad (6.18)$$

After inserting this result into (6.13), the bound on false dismissal probability becomes

$$1-\beta \leq \frac{1}{\alpha_0} e^{nr^*(g_1)} \tag{6.19}$$

Thus, the bound on this error decreases exponentially with the number of data observations.

Two detectors, g and h, may be compared by computing their *relative efficiency*, where $\mathrm{RE}_{g,h} = \frac{n_h(\alpha_0,\beta)}{n_g(\alpha_0,\beta)}$, the ratio of the number of observations in the respective detectors operating with false alarm rate no greater than α_0 and probability of correct detection at least β. While $r^*(g)$ does not allow computation of the exact value of β, it does allow computation of a bound on $1-\beta$.

Proposition 3. Suppose two memoryless detector nonlinearities g and h operating on iid distributed observations each use n_g and n_h data observations, respectively, and $\frac{n_h}{n_g} = \frac{r^*(g_1)}{r^*(h_1)}$. Then

$$\frac{1}{\alpha_0} e^{n_g r^*(g_1)} = \frac{1}{\alpha_0} e^{n_h r^*(h_1)}$$

Proof. The proof follows from direct computation.■

The quantity $\frac{r^*(g_1)}{r^*(h_1)}$ may be designated as the *relative bound efficiency* of detector g relative to detector h. Thus

$$\mathrm{RBE}_{g,h} = \frac{r^*(g_1)}{r^*(h_1)} \tag{6.20}$$

is a measure of the relative rates of convergence in the false dismissal probability of two detectors operating with equal false alarm rates. Alternatively, it may be considered as a measure of the ratio of the number of samples needed in each detector to obtain equal bounds on the false dismissal probability for equal false alarm rates. Note that (6.20) extends easily by replacing $r^*(g_1)$ with $r^*(g)$. Thus, the RBE of two detectors may be compared for non-iid noises, and detectors with memory.

A related measure of efficiency is the Chernoff asymptotic relative efficiency [2,11], or $ARE^c_{g,h}$, defined as

$$ARE^c_{g,h} = \frac{\min\left[\min_u M_0(u;g), \min_u M_1(u;g)\right]}{\min\left[\min_u M_0(u;h), \min_u M_1(u;h)\right]}$$

The proposed measure RBE differs from ARE^c in that RBE measures the relative rates of convergence of $1-\beta$ under equal false alarm rates for the two detectors.

Detection of a Known Constant Signal

An often discussed special case is the problem of detecting the presence or absence of a known constant signal in the presence of an additive iid noise. When the signal is positive, this problem is sometimes is known as the *shift-to-the-right* problem. The univariate noise densities under the respective hypotheses become

$$f_0(x) = f(x)$$
$$f_1(x) = f(x-\theta s)$$

where θ is the known constant signal amplitude, and f is the univariate density of the additive noise. For convenience, and without loss of

generality, we shall hereafter assume $s = 1$. A common situation is for the noise to have zero mean, and for the density to be symmetric about the mean. The optimal detector will then be odd-symmetric about the point $\theta/2$. Under these conditions, we have

Proposition 4. If

$$f_1(x) = f_0(x-\theta) \tag{6.21}$$

$$f_0(x) = f_0(-x) \tag{6.22}$$

$$g\left[\frac{\theta}{2}-x\right] = -g\left[\frac{\theta}{2}+x\right] \tag{6.23}$$

then $M_1(u;g) = M_0(u;g)$.

Proof. The proof begins with the definition

$$M_1(u;g) = \ln \int_{-\infty}^{\infty} e^{-ug(x)} f_1(x)dx$$

Applying (6.21) yields

$$M_1(u;g) = \ln \int_{-\infty}^{\infty} e^{-ug(\theta-x)} f_0(x)dx$$

$$= \ln \int_{-\infty}^{\infty} e^{-ug(\theta/2+(x+\theta/2))} f_0(x)dx$$

and applying (6.23) gives

$$M_1(u;g) = \ln -\int_{\infty}^{-\infty} e^{ug(x)} f_0(-x)dx$$

Finally application of (6.22) yields the desired result

$$M_1(u;g) = M_0(u;g)$$

■

It is possible to show in the shift-to-the-right problem that, as the signal vanishes, the quantity $\mathrm{RBE}_{g,h}$ for two detectors approaches asymptotically the value of $\mathrm{ARE}_{g,h}$.

Theorem 4. In the shift-to-the-right problem, with iid noise and detectors $g(\mathbf{x}) = \sum_{i=1}^{n} g(x_i;\theta)$ and $h(\mathbf{x}) = \sum_{i=1}^{n} h(x_i;\theta)$ that are odd-symmetric about $\theta/2$, with test structure (6.6), and with test thresholds $\mathrm{E}_0 g \le t_g \le \mathrm{E}_1 g$ and $\mathrm{E}_0 h \le t_h \le \mathrm{E}_1 h$, respectively, let the false alarm rate be equal in both tests. Then

$$\lim_{\theta \to 0} \mathrm{RBE}_{g,h} = \mathrm{ARE}_{g,h}$$

Proof. The power of the test using g is

$$\begin{aligned} \beta_g &= \mathrm{Prob}\left[g(\mathbf{x}) \ge n t_g \mid \mathrm{H}_1 \right] \\ &= \mathrm{Prob}\left[h(\mathbf{x}) < n t_g \mid \mathrm{H}_0 \right] \end{aligned}$$

and similarly for β_h. By application of Chernoff's theorem [8, 11] it may be shown after some simple algebra that

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \ln \mathrm{Prob}\left[g(\mathbf{x}) \ge n t_g \mid \mathrm{H}_1 \right] \\ = \min_u \left[-u t_g + M_1(u;-g) \right] \end{aligned} \tag{6.24}$$

and that

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \ln \mathrm{Prob}\left[g(\mathbf{x}) < n t_g \mid \mathrm{H}_0 \right] \\ = \min_u \left[u t_g + M_0(u;-g) \right] \end{aligned} \tag{6.25}$$

independently of the value of θ. By Proposition 4, $M_0(u;-g) = M_1(u;-g)$. Therefore

$$\min_u \left[-ut_g + M_1(u;-g)\right] \qquad (6.26)$$
$$= \tfrac{1}{2}\min_u \left[M_0(u;-g-t_g) + M_1(u;-g+t_g)\right]$$

and

$$\min_u \left[ut_g + M_0(u;-g)\right] \qquad (6.27)$$
$$= \tfrac{1}{2}\min_u \left[M_0(u;-g+t_g) + M_1(u;-g-t_g)\right]$$

However, (6.24) and (6.25) imply that (6.26) and (6.27) must be equal. As a consequence of Proposition 1, they are equal to $\tfrac{1}{2}r^*(g)$.

Following Capon [9], let $\{\theta_k\}$ be a sequence of signals such that $\lim_{k\to\infty}\theta_k = 0$, and let the sequences $\{n_{g;k}\}$ and $\{n_{h;k}\}$ be two increasing sequences of integers such that

$$0 \neq \lim_{k\to\infty}\beta_g(\theta_k, n_{g;k}) = \lim_{k\to\infty}\beta_h(\theta_k, n_{h;k}) \neq 1 \qquad (6.28)$$

Since the nonlinearities g and h are functions of θ, we will denote the sequences of nonlinearities dependent on $\{\theta_k\}$ as g_k and h_k, respectively. Then

$$\lim_{k\to\infty}\ln\beta_g(\theta_k, n_{g;k}) = \lim_{k\to\infty}\ln \mathrm{Prob}[g_k(\mathbf{x}) \geq nt_g \mid \mathrm{H}_1]$$

and similarly for detector h_k. The ratio of false dismissal probabilities for the two detectors may be written as

$$\frac{n_{h;k}\ln\beta_g(\theta_k, n_{g;k})}{n_{g;k}\ln\beta_h(\theta_k, n_{h;k})} = \frac{n_{h;k}\ln \mathrm{Prob}\left[g(\mathbf{x}) \geq n_{g;k}t_g \mid \mathrm{H}_1\right]}{n_{g;k}\ln \mathrm{Prob}\left[g(\mathbf{x}) \geq n_{h;k}t_h \mid \mathrm{H}_1\right]} \qquad (6.29)$$

By previous arguments, it follows that in the limit the right

side of (6.29) becomes the ratio

$$\lim_{k \to \infty} \frac{r^*(g_k)}{r^*(h_k)} \tag{6.30}$$

which by the definition (6.20) of RBE is the quantity $\lim_{k \to \infty} \mathrm{RBE}_{g_k, h_k}$. Condition (6.28) assures that in the limit the powers of detectors g_k and h_k are equal, which reduces the left side of (6.29) to the definition of asymptotic relative efficiency

$$\lim_{k \to \infty} \frac{n_{h;k}}{n_{g;k}} = \mathrm{ARE}_{g,h}$$

The conclusion then is that as $\theta \to 0$, the quantities $\mathrm{RBE}_{g,h}$ and $\mathrm{ARE}_{g,h}$ are asymptotically equivalent. Note that in the limit, nonlinearities g_k and h_k are odd-symmetric about the origin. ■

Numerical Examples

In this section, the performance index is calculated and compared for three different detector structures in two different noise environments for the shift-to-the-right problem. The objective is to decide between

$$\begin{aligned} &\mathrm{H}_0\colon\ x_i \sim f(x_i) \\ &\mathrm{H}_1\colon\ x_i \sim f(x_i - \theta) \end{aligned} \qquad \text{for } i = 1, \ldots, n \text{ and } \theta > 0$$

using a test

$$\sum_{i=1}^{n} g(x_i) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} t$$

The three detector nonlinearities which will be examined are the *linear detector*

$$g_{ld}(x) = \theta(x-\theta/2) \tag{6.31}$$

the *sign detector*

$$g_{sd}(x) = \mathrm{sgn}(x-\theta/2) \tag{6.32}$$

and the *amplifier limiter*

$$g_{al}(x) = \begin{cases} -\theta\sqrt{2} & \text{for } -\infty \le x < 0 \\ 2\sqrt{2}(x-\theta/2) & \text{for } 0 \le x < \theta \\ \theta\sqrt{2} & \text{for } \theta \le x < \infty \end{cases} \tag{6.33}$$

The two densities which will be used are the *Gaussian* density

$$f_G(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{6.34}$$

and the *Laplace* density

$$f_L(x) = \frac{\sqrt{2}}{2} e^{-\sqrt{2}|x|} \tag{6.35}$$

The three detectors are illustrated in Figures 6.1 - 6.3. Note that $g_{ld}(x)$ is the Neyman-Pearson optimal nonlinearity for $f = f_G$, and g_{sd} is the NP optimal nonlinearity for $f = f_L$.

The methods of this chapter may be applied to calculate $r^*(g)$, and the RBE of various pairs of detectors under the two noise environments. Appendix 6.1 gives the formulation of $r(u;g;f)$ for all six combinations of detector nonlinearities and densities. Here, f appears as an argument of r to emphasize the dependence of $r^*(g;f)$ on a single univariate density. For some combinations of nonlinearities and densities, r^* or u^* is given,

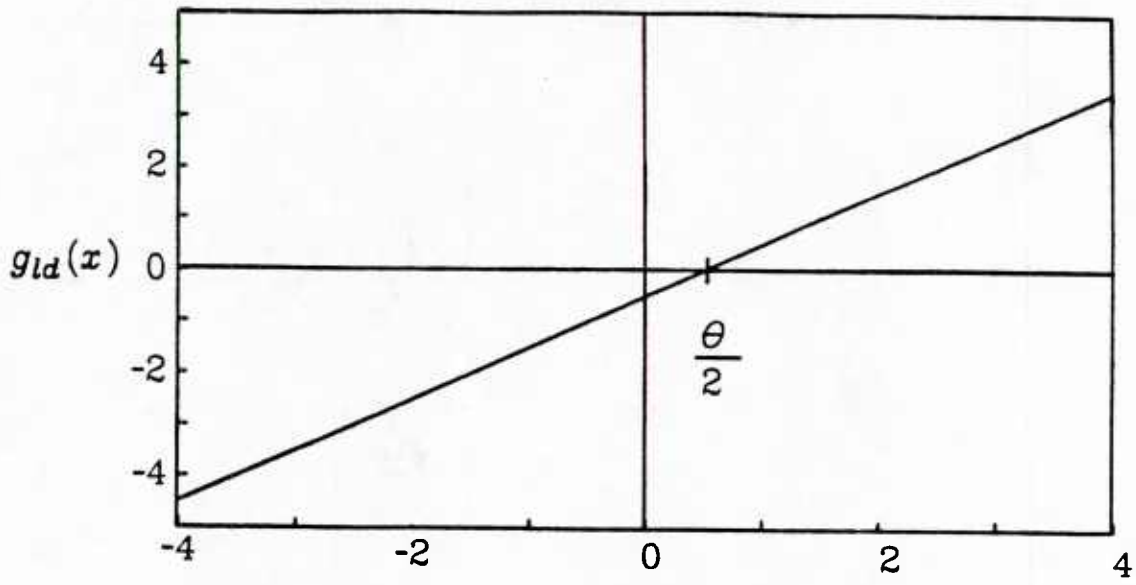

Fig. 6.1. The linear detection processor g_{ld} for $\theta = 1$.

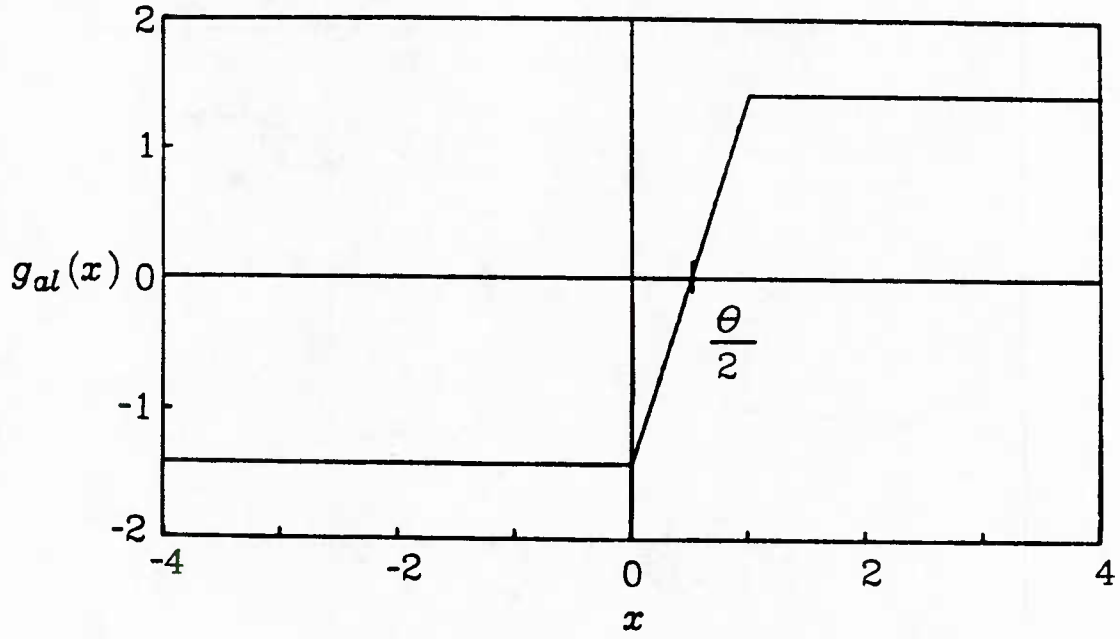

Fig. 6.2. The amplifier limiter detector nonlinearity g_{al} for $\theta = 1$.

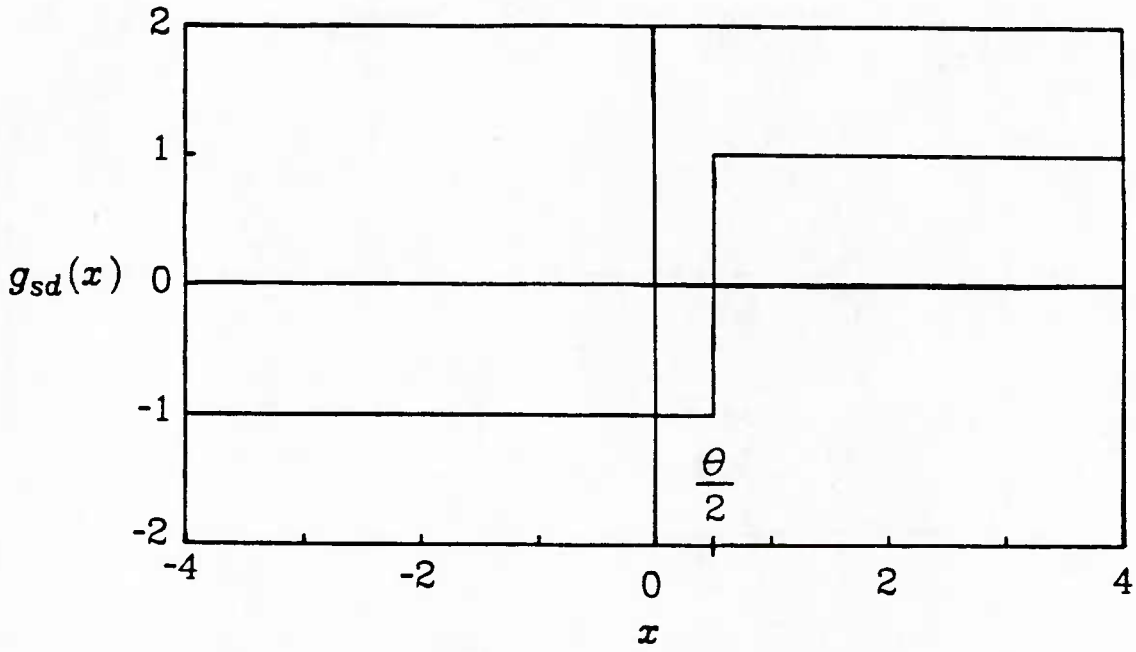

Fig. 6.3. The sign detector nonlinearity g_{sd} for $\theta = 1$.

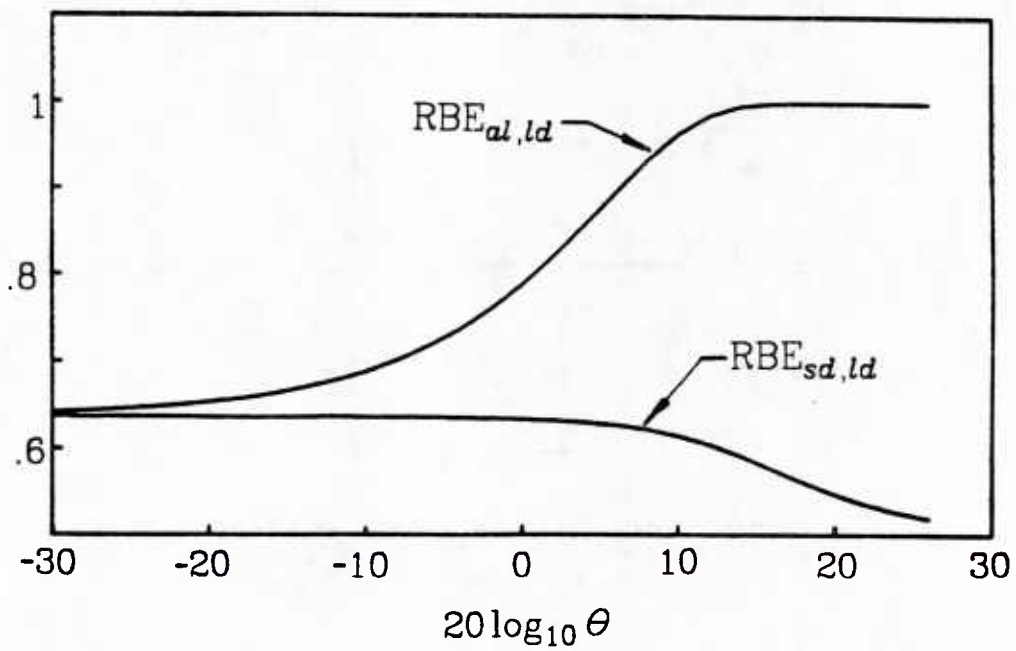

Fig. 6.4. Performance comparison of the amplifier limiter and the sign detector relative to the linear detector in Gaussian noise.

but for others this value must be found through numerical methods.

Figures 6.4 and 6.5 present the RBE of the detector pairs under Gaussian and Laplace noise assumptions, respectively. Since both densities were defined with unit variance, the horizontal axis of the plots is also a measure of the *signal-to-noise-ratio* (SNR). The nonlinearities are parameterized as a function of θ; thus as θ becomes small the shape of g_{al} and g_{sd} become nearly identical relative to a fixed observation scale.

As predicted in Theorem 4, $\mathrm{RBE}_{al,ld}$ and $\mathrm{RBE}_{sd,ld}$ asymptotically approach $\mathrm{ARE}_{sd,ld}$ for small θ. When the SNR, (equivalently, θ), becomes large, $\mathrm{RBE}_{al,ld}$ approaches unity for both densities, implying that under this condition the amplifier limiter and the linear detector have the same efficiency. Also, $\mathrm{RBE}_{sd,ld}$ converges to ½, as shown in Appendix 6.2.

For comparison, the ARE of various detector pairs may also be calculated as a function of the parameter θ. For the purpose of calculating efficacy, it is assumed that the nonlinearities are symmetric about zero instead of $\theta/2$. Therefore, in Appendix 6.1 the efficacy is given for the shifted nonlinearities $g(x+\theta/2)$. Figures 6.6 - 6.11 compare RBE and ARE for pairs of detectors under the different noise assumptions.

All six of the figures further emphasize the convergence of $\mathrm{ARE}_{al,ld}$, $\mathrm{RBE}_{al,ld}$, and $\mathrm{RBE}_{sd,ld}$ for small θ. The performance of the amplifier limiter and the linear detector are approximately equivalent for high SNR, as shown by both ARE and RBE. Notice that, while ARE predicts a constant performance level for the sign detector, Figures 6.8 - 6.11 emphasize that the linear detector or amplifier limiter may well outper-

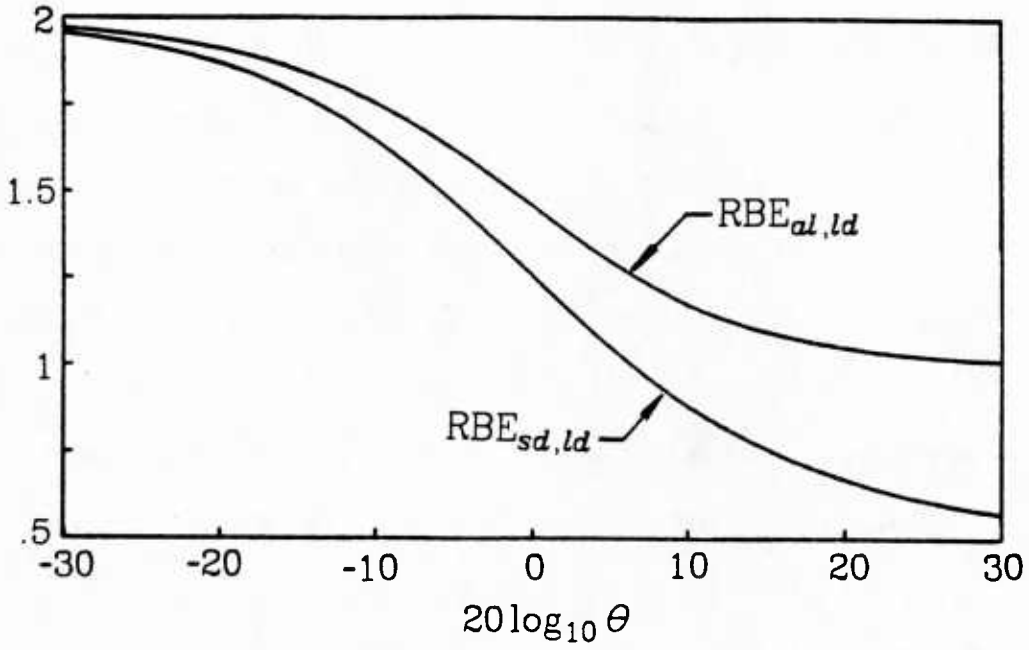

Fig. 6.5. Performance comparison of the amplifier limiter and the sign detector relative to the linear detector in Laplace noise.

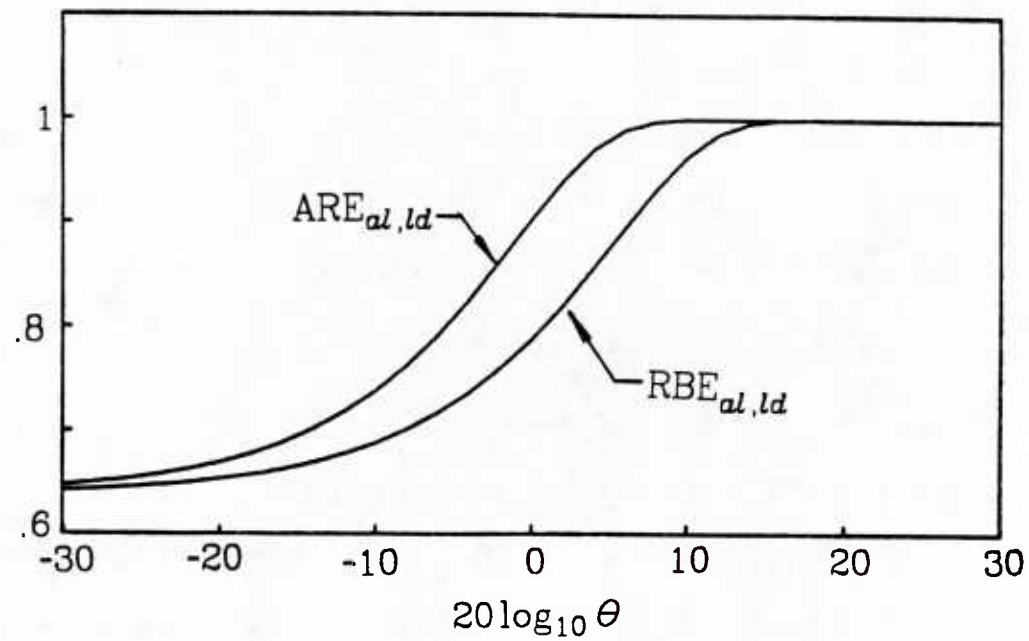

Fig. 6.6. Comparison of ARE and RBE of the amplifier limiter relative to the linear detector in Gaussian noise.

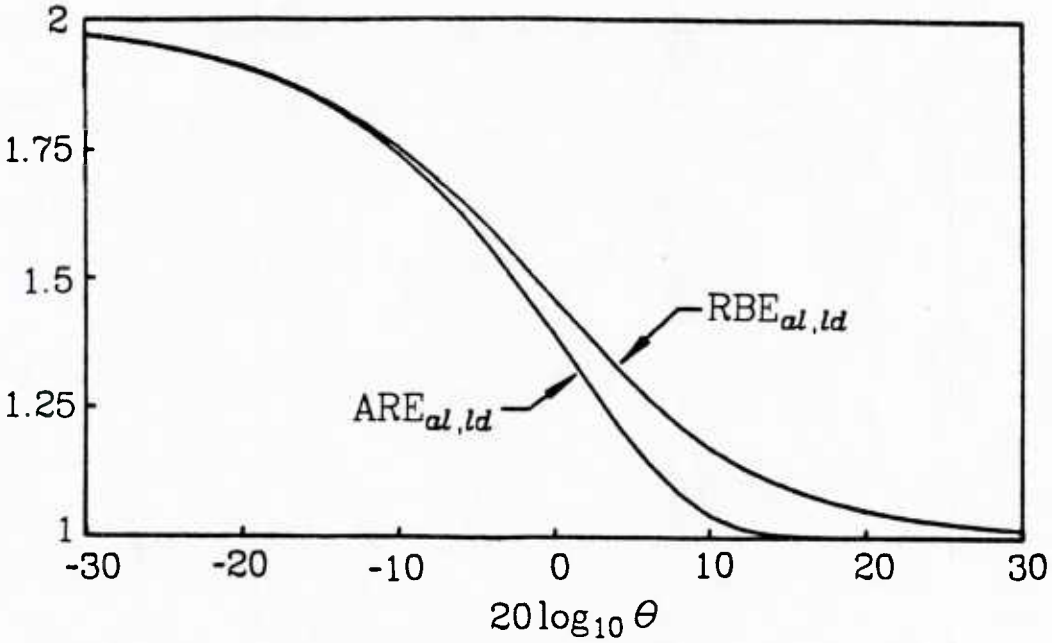

Fig. 6.7. Comparison of ARE and RBE of the amplifier limiter relative to the linear detector in Laplace noise.

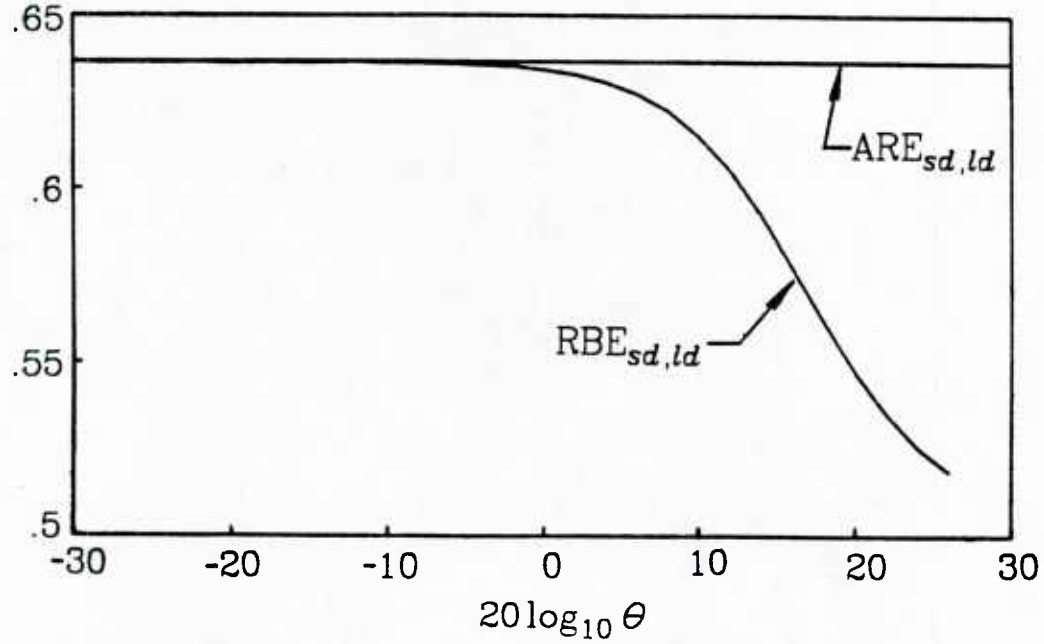

Fig. 6.8. Comparison of ARE and RBE of the sign detector relative to the linear detector in Gaussian noise.

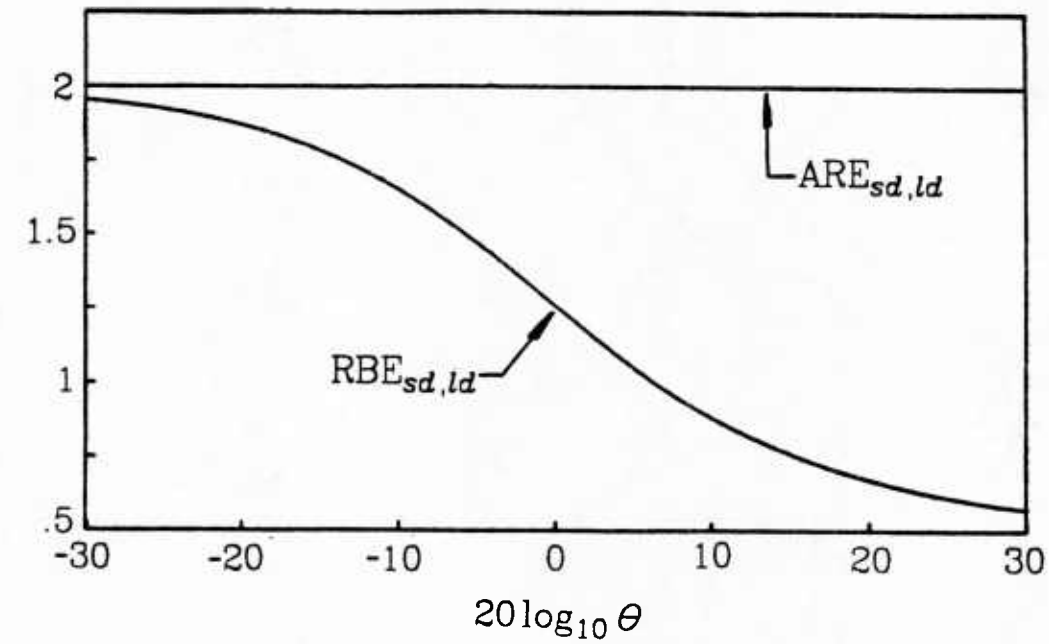

Fig. 6.9. Comparison of ARE and RBE of the sign detector relative to the linear detector in Laplace noise.

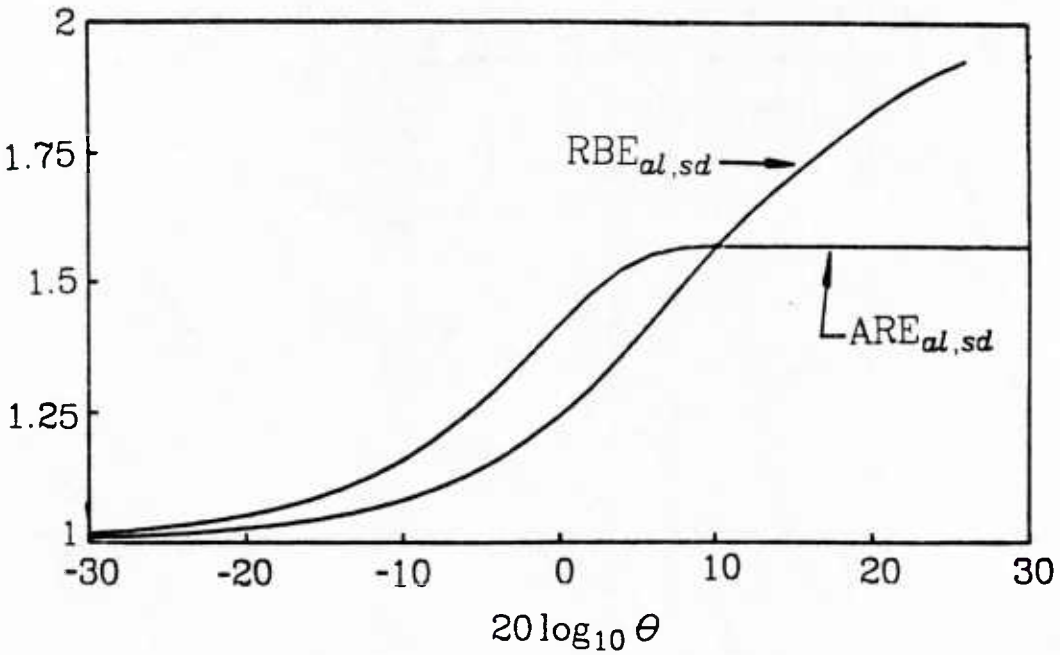

Fig. 6.10. Comparison of ARE and RBE of the amplifier limiter relative to the sign detector in Gaussian noise.

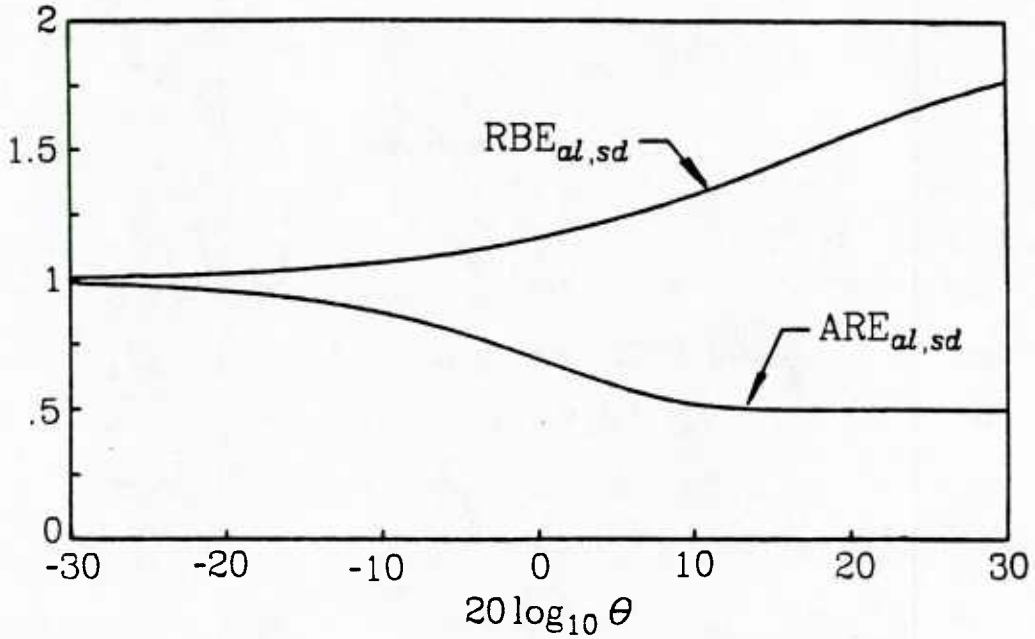

Fig. 6.11. Comparison of ARE and RBE of the amplifier limiter relative to the sign detector in Laplace noise.

form the sign detector for moderate to high SNR. Michalsky, Wise and Poor [12] studied the convergence of relative efficiency to asymptotic relative efficiency in the finite sample size detector and observed similar difficulties with ARE. They also found that in certain cases relative efficiency may produce a different ranking of detector performance than would asymptotic relative efficiency

3. Conclusion

Some properties of a functional $r^*(g)$ were developed in this chapter, and it was shown that $r^*(g)$ is a performance measure which may be potentially useful for studying the performance of finite sample size detectors. In this regard, it may be considered as a figure of merit, or a performance index for a detector. As was demonstrated, $r^*(g)$ is a quantity which may be used to form an exponential bound on $1-\beta$. Thus, the smaller the value $r^*(g)$, the smaller the bound on false dismissal probability. Given a pair of hypotheses, $r^*(g)$ may be used to rank competing alternative structures.

A disadvantage of bounding methods is that it is not clear that comparing and ordering systems by a performance bound corresponds exactly to an ordering of the systems by their error probabilities. Indeed, we resort to a bounding method precisely because we are unable to calculate, (and hence, order by) the error probabilities of the alternative systems. A bound is useful for comparing the relative merits of systems, though, for a bound guarantees a certain minimum performance level. As a result, it is reasonable to say that the tightest bound corresponds in some sense to the "best" system.

Of particular utility is the ratio of respective $r^*(g)$ indices for two competing detectors. This ratio was denoted as relative bound efficiency, or RBE, as was shown to be asymptotically equivalent to ARE as the signal to noise ratio vanishes. For finite SNR, however, the behavior of RBE diverges from ARE and follows more closely intuition about the relative efficiency of several common finite sample size detectors.

Appendix 6.1

In the previously stated conditions of the shift-to-the-right problem $r(u;g) = 2M_0$. Therefore, for the six combinations of nonlinearities and noise densities, an expression for M_0 will be given, and when a simple form can be found, an expression for $r^*(u;g)$. For comparison, the expression for efficacy of the shifted (zero-centered) nonlinearities are also given. The cumulative distribution of the Gaussian density is written here as $\Phi(x)$.

1. Gaussian density, linear detector

$$r(u;g_{ld};f_G) = u(u-1)\theta^2 \tag{A6.1}$$

$$r^*(g_{ld};f_G) = -\frac{\theta^2}{4} \tag{A6.2}$$

$$\eta_G(g_{ld}(x+\theta/2;\theta)) = 1 \tag{A6.3}$$

2. Gaussian density, sign detector

$$r(u;g_{sd};f_G) = 2\ln\left[e^{-u}\Phi(\theta/2)+e^{u}\Phi(-\theta/2)\right] \tag{A6.4}$$

$$r^*(g_{sd};f_G) = 2\ln 2+\ln\Phi(\theta/2)+\ln\Phi(-\theta/2) \tag{A6.5}$$

$$\eta_G(g_{sd}(x+\theta/2;\theta)) = \frac{2}{\pi} \tag{A6.6}$$

3. Gaussian density, amplifier limiter

$$r(u;g_{al};f_G) = \tag{A6.7}$$
$$2\left[\Phi(\theta-2\sqrt{2}u)-\Phi(-2\sqrt{2}u)\right]\exp\left[-(\sqrt{2}u\theta-4u^2)\right]$$

$$\eta_G(g_{al}(x+\theta/2);\theta)) = \frac{\left[1-\Phi(-\theta/2)\right]^2}{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{\theta^2}{4},\frac{3}{2}\right)+\frac{\theta^2}{2}\Phi(-\theta/2)} \quad \text{(A6.8)}$$

Here, Γ is the incomplete gamma function

$$\Gamma(x,y) = \int_0^x e^{-\tau}\tau^{y-1}d\tau$$

4. Laplace density, linear detector

$$r(u;g_{ld};f_L) = 2\ln 2 - u\theta^2 - 2\ln(2-u^2\theta^2) \quad \text{(A6.9)}$$

$$u^* = 2\frac{-1+\sqrt{1+\theta^2/2}}{\theta^2} \quad \text{(A6.10)}$$

$$\eta_L(g_{ld}(x+\theta/2;\theta)) = 1 \quad \text{(A6.11)}$$

5. Laplace density, sign detector

$$r(u;g_{sd};f_L) = 2\ln\left[e^{-u} + \frac{e^{-\theta\sqrt{2}/2}}{2}(e^u - e^{-u})\right] \quad \text{(A6.12)}$$

$$r^*(g_{sd};f_L) = \ln 2 - \frac{\theta\sqrt{2}}{2} + \ln\left[1-\frac{e^{-\sqrt{2}\theta/2}}{2}\right] \quad \text{(A6.13)}$$

$$\eta_L(g_{sd}(x+\theta/2;\theta)) = 2 \quad \text{(A6.14)}$$

6. Laplace density, amplifier limiter

$$r(u;g_{al};f_L) = \quad \text{(A6.15)}$$

$$2\ln\left[e^{-u\theta\sqrt{2}} + e^{(u-1)\theta\sqrt{2}} + \frac{e^{-u\theta\sqrt{2}}}{\sqrt{2}(1-2u)}(1-e^{(2u-1)\theta\sqrt{2}})\right] - \ln 2$$

$$r^*(g_{al};f_L) = 2\ln\left[e^{-\theta\sqrt{2}}(2+\theta)\right] - 2\ln 2 \qquad \text{(A6.16)}$$

$$\eta_L(g_{al}(x+\theta/2;\theta)) = \frac{\left[1-e^{-\sqrt{2}\theta^2/2}\right]^2}{1-\frac{e^{-\sqrt{2}\theta/2}}{2}(\theta^2+\theta\sqrt{2}+2)} \qquad \text{(A6.20)}$$

Appendix 6.2

Here it is demonstrated that $\mathrm{RBE}_{sd,ld}$ approaches asymptotically the value ½ for the shift-to-the-right problem and iid observations. To begin, let $p = F(\theta/2)$, where F is the cumulative distribution function of the noise density. Then

$$M_0(u;g_{sd}) = \ln\left[e^{-u}p + e^{u}(1-p)\right] \tag{A6.18}$$

and it is easy to show that $u^* = \tfrac{1}{2}\ln\frac{p}{1-p}$. Using this value of u^*,

$$r_1^*(u^*,g_{sd}) = \ln p + \ln(1-p) + 2\ln 2 \tag{A6.19}$$

for, as a consequence of Proposition 4, $r_1 = 2M_0$.

It may be shown [10] through a saddlepoint expansion approach that

$$\ln(1-p) = M_0(u^*,g_{ld}) + \varepsilon(\ln\theta^{-\frac{1}{2}}) \tag{A6.20}$$

where ε represents the approximation error, of order $\ln\theta^{-\frac{1}{2}}$. Using (A6.19) and (A6.20) the ratio $RBE_{sd,ld}$ may be written as

$$\frac{r_1^*(u^*,g_{sd},f)}{r_1^*(u^*,g_{ld},f)} = \frac{\ln p + \ln(1-p) + 2\ln 2}{2\ln(1-p) + 2\varepsilon(\theta^{-\frac{1}{2}})} \tag{A6.21}$$

Finally, after noting that $\lim_{\theta\to\infty} p = 1$ and applying L'Hôpital's Rule to (A6.21), we conclude that

$$\lim_{\theta\to\infty} \mathrm{RBE}_{sd,ld} = \tfrac{1}{2} \tag{A6.22}$$

References

[1] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, Inc: San Francisco, 1977.

[2] H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations", *Ann. Math. Stat.*, vol. 23, pp. 493-507, 1952.

[3] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection", *IEEE Trans. Commun. Tech.*, vol. COM-15, no. 1, pp. 52-60, Feb. 1967.

[4] R.E. Blahut, "Hypothesis Testing and Information Theory", *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 405-417, July 1974.

[5] D. Kazakos, "Signal Detection under Mismatch", *IEEE Trans. Inform. Theory*, vol. IT-28, no. 4, pp. 681-684, July 1982.

[6] D. Kazakos, "Statistical Discrimination using Inaccurate Models", *IEEE Trans. Inform. Theory*, vol. IT-28, no. 5, Part I, pp. 720-728, Sept. 1982.

[7] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons: New York, 1971.

[8] P. Billingsley, *Probability and Measure*, John Wiley and Sons: New York, 1979.

[9] J. Capon, "On the Asymptotic Relative Efficiency of Locally Optimum Detectors", *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67-71, April 1961.

[10] F.W.J. Olver, *Asymptotics and Special Functions*, Academic Press: New York, 1974.

[11] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons: New York, 1980.

[12] D.L. Michalsky, G.L. Wise, and H.V. Poor, "A Relative Efficiency Study of some Popular Detectors", *J. of Franklin Inst.*, vol. 313, pp. 135-148, March 1982.

[13] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, Inc.: New York, 1969.

7

Conclusion

In this chapter, the main contributions of the dissertation are reviewed, and some suggestions for extending this research are made.

1. Review and Suggestions

Chapter 2

Detection procedures and noise models were highlighted in Chapter 2, and the failings of the classical Gaussian noise assumption were noted. As an alternative model, a definition of non-Gaussian noise density was given, and several commonly used non-Gaussian noise models were exhibited.

The critical feature of non-Gaussian noise as defined here is the fact that the density is much heavier tailed than the Gaussian density. In this type of detection environment it is important to reduce the influence of the very large observations; even just a few impulses, or outliers, can

seriously disturb detector performance.

Finally, in light of the work in robust statistics, and the work in optimal detection in non-Gaussian noise, it was proposed to examine simple adaptive detectors which are useful when only a very loose characterization of the noise statistics is available.

At the end of the chapter, some Arctic under-ice noise data was discussed. It would be interesting to study its characteristics further, particularly examining its distribution and dependency structure. Does the data fit any commonly used models?

Chapter 3

The following conjecture was proposed and exploited successfully: Suppose some generic detector nonlinearity with a roughly linear region near the origin is chosen that allows freedom in selection of the nonlinearity tail behavior. Then, it should be possible to make measurements on the observed noise and adjust the nonlinearity tails appropriately.

Two alternatives techniques were proposed and studied: the tail matching method, giving g_{tm}, and the efficacy maximizing procedure, leading to the piecewise linear processor g_{2l}. In the examples, both methods were able to achieve high levels of performance relative to the optimal structure, even though both nonlinearities were *ad hoc* proposals and only very simple measurements of the noise density were used to drive adaptation. When simulated with the physical noise data, both detectors appeared to have improved performance relative to the linear

detector.

The conclusion to be drawn from the chapter is that, when choosing the form for a nonlinearity, it is not critical to find the exactly optimal structure. Rather, it is possible to achieve nearly optimal results using quite simple structures, provided that there is enough freedom to adapt the structure to the particular noise density of interest. As noted in Chapter 2, the specification of a particular class of non-Gaussian densities leads to generic specification of the class of suitable approximate nonlinearities.

It would be worthwhile to investigate further certain properties of suboptimal detectors. For instance, the performance of a suboptimal nonlinearity is sometimes less sensitive to changes in the noise environment than the optimal nonlinearity. What causes this property, and how may it best be employed? Can other methods besides Huber's min-max approach produce robust detector nonlinearities?

Chapter 4

When a nominal background noise is contaminated by bursts of impulsive noise, it was shown that it is possible to design a structure which recognizes the bursts, and then uses this information to adapt the detector rapidly. The structure was developed in two parts: one part was a time varying detector which switched between two nonlinearities, and the other part was a nonparametric noise burst detector utilizing a median filter.

Under one reasonable and realistic set of assumptions, it was demon-

strated that the switched burst detector can outperform any fixed detector structure.

One problem mentioned in the chapter and worthy of further attention is analysis of the switched burst detector algorithm when a statistical model for the noise burst lengths is available. Also, given a statistical description of the burst run lengths, how may the nonparametric burst detector be improved? Probably, this knowledge would lead to a threshold test where the threshold varied as a function of the number of observations since the last state transition was encountered.

Another important area to be investigated is the use of alternatives to linear detectors during the impulsive noise modes. Would any performance advantage due to the use of robust nonlinearities outweigh the loss of simplicity when the low gain linear alternative is replaced?

Chapter 5

In Chapter 5 the equivalence between efficacy maximizing procedures and minimum mean square approximation of the true locally optimum nonlinearity is demonstrated. In particular, the results lend substance to some loose ideas about what constitutes a "good" approximation: it is important to match the optimal nonlinearity closely in the regions where an observation is highly probable, while a rough approximation is sufficient in low probability regions such as the density tails. Moreover, once an approximation is fairly "close" to the true nonlinearity, further refinements lead to little performance improvement. This is not to say that any nonlinearity tail behavior will suffice; the mean square error between a linear processor and blanking nonlinearity tails can be

great, despite weighting by a relatively small probability mass, and it is known that the linear processor performs poorly in heavy-tailed noises.

This chapter provides a distance measure between nonlinearities. Is it possible to find a min-max robust suboptimal nonlinearity using this tool? Another useful extension of this work might include examination of nonlinearity approximation procedures in the dependent noise case.

Chapter 6

A performance index r^*, useful for comparing detectors operating with equal false alarm rates, was developed in Chapter 6. It was shown that the ratio of performance indices for two detectors is a useful indicator of their relative performance under non-zero signal to noise ratios. Further, this ratio, the proposed measure of relative bound efficiency, approaches the measure of asymptotic relative efficiency as the signal vanishes.

It would be worthwhile to examine r^* and relative bound efficiency further. For instance, it would be interesting to examine their use in dependent noise. There are other open points: how does relative bound efficiency compare to relative efficiency? How tight is the bound on performance using r^*? Is it possible to find r^* directly and circumvent the proposed minimization procedure?

2. Conclusion

The underlying goal of this study was to consider the signal detection problem in the case of incomplete knowledge of the non-Gaussian noise environment. In striving towards this goal, work was presented ranging

from simulations using physical noise data to theoretical analysis.

The theoretical results of this thesis may be useful tools in the continued study of nearly optimal detectors. The proposals for detector structures presented here are not definitive; however, they do confirm some ideas about useful approaches to this problem, and point out possible directions for further research.

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

FEBRUARY 1982

	Copies
Statistics and Probability Program (Code 411(SP)) Office of Naval Research Arlington, VA 22217	3
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Commanding Officer Office of Naval Research Eastern/Central Regional Office Attn: Director for Science Barnes Building 495 Summer Street Boston, MA 02210	1
Commanding Officer Office of Naval Research Western Regional Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101	1
U. S. ONR Liaison Office - Far East Attn: Scientific Director APO San Francisco 96503	1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G. H. Gleissner Bethesda, Maryland 20084	1
Commandant of the Marine Coprs (Code AX) Attn: Dr. A. L. Slafkosky Scientific Advisor Washington, DC 20380	1
Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522	1
U. S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706	1
Director National Security Agency Attn: R51, Dr. Maar Fort Meade, MD 20755	1
ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002	1
ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403	1
Library, Code 1424 Naval Postgraduate School Monterey, CA 93940	1
Technical Information Division Naval Research Laboratory Washington, DC 20375	1
OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301	1

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390

	Copies		Copies
Mr. F. R. Del Priori Code 224 Operational Test and Evaluation Force (OPTEVFOR) Norfolk, VA 23511	1		